WOODHOUSE EXHIBIT 10

UNREDACTED VERSION OF EXHIBIT A [DKT. 357-1]

FILED UNDER SEAL

Joseph R. Saveri (State Bar No. 130064)
Cadio Zirpoli (State Bar No. 179108)
Christopher K.L. Young (State Bar No. 318371)
Holden Benon (State Bar No. 325847)
Kathleen J. McMahon (State Bar No. 340007)
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, CA 94108
Telephone: (415) 500-6800
Facsimile: (415) 395-9940
Email: jsaveri@saverilawfirm.com
              czirpoli@saverilawfirm.com
              cyoung@saverilawfirm.com
              hbenon@saverilawfirm.com
              kmcmahon@saverilawfirm.com

Matthew Butterick (State Bar No. 250953)
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Telephone: (323) 968-2632
Facsimile: (415) 395-9940
Email: mb@buttericklaw.com

Counsel for Individual and Representative Plaintiffs
and the Proposed Class

[Additional Counsel Listed on Signature Page]

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

Richard Kadrey, Sarah Silverman, Christopher Golden, Michael Chabon, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Ayelet Waldman, and Jacqueline Woodson, *Individual and Representative Plaintiffs,* v. Meta Platforms, Inc., a Delaware corporation; *Defendant.*	Case No. 3:23-cv-03417-VC **PLAINTIFFS' FIRST SET OF INTERROGATORIES TO DEFENDANT META**

In accordance with Federal Rules of Civil Procedure Rules 26 and 33, Plaintiffs, by and through their undersigned attorneys, hereby request that Defendant Meta respond to the following First Set of Interrogatories separately and fully, in writing and under oath, within thirty (30) days of service hereof, in accordance with the following definitions and instructions.

DEFINITIONS

As used herein, the following words, terms, and phrases—whether singular or plural, or in an alternate verb tense—shall have the meanings ascribed below. Defined terms may not be capitalized or made uppercase. The given definitions apply even if a term in question is not capitalized or made uppercase. No waiver of a definition is implied by the use of a defined term in a non-capitalized or lowercase form:

1. "Agreements" means any oral or written contract, arrangement or understanding, whether formal or informal, between two or more Persons, including all drafts, versions, modifications, amendments, attachments, exhibits, and appendices thereof.

2. "All," "Or," and "And" should be understood to include and encompass "any"; "or" should be understood to include and encompass "and"; and "and" should be understood to include and encompass "or."

3. "Communications" means oral or written communications of any kind, communicated directly or indirectly, Including, without limitation inquiries, complaints, discussions, conversations, negotiations, agreements, meetings, interviews, telephone conversations, letters, correspondences, memoranda, notes, telegrams, facsimiles, electronic mail (e-mail) messages and attachments, instant or direct messages (Including SMS messages, text messages, Apple iMessages, Slack messages, Teams messages), memoranda, documents, writings, or other forms of communications. The term "Communications" Includes instances where one party disseminates information that the other party receives but does not respond to.

4. "Complaint" refers to the operative complaint at the time documents are produced in response to these requests. At the time of service, the currently operative Complaint is Plaintiffs' First Amended Complaint. ECF No. 64.

5. "Concerning" refers to and includes "constituting," "evidencing," "supporting,"

1 "regarding," "mentioning," "reflecting," "concerning," "relating to," "referring to," "pertaining

2 to," "alluding to," "responding to," "proving," "discussing," "assessing," "disproving,"

3 "connected with," "commenting on," "about," "showing," "describing," and/or logically or

4 factually dealing with the matter described in the request in which the term appears.

5   6.   "Defendant" means Defendant Meta Platforms, Inc.

6   7.   "Document" is used in its broadest sense allowed by Rule 34(a) of the Federal Rules

7 of Civil Procedure and includes, but is not limited to, any writings, drawings, graphs, handwriting,

8 typewriting, printing, photostatting, photographing, photocopying, transmitting by electronic mail

9 or facsimile, and every other means of recording upon any tangible thing, any form of

10 communication or representation, Including letters, words, pictures, sounds, or symbols, or

11 combinations thereof, and any record thereby created, regardless of the manner in which the record

12 has been stored.

13   This Includes:

14   •   The originals, drafts and All non-identical copies thereof, whether different from the

15   original by reason of any notation made on such copies or otherwise;

16   •   Booklets, brochures, pamphlets, circulars, notices, periodicals, papers, contracts,

17   agreements, photographs, minutes, memoranda, messages, appraisals, analyses,

18   reports, financial calculations and representations, invoices, accounting and diary

19   entries, inventory sheets, diaries, appointment books or calendars, teletypes,

20   telefaxes, thermafaxes, ledgers, trial balances, correspondence, telegrams, press

21   releases, advertisements, notes, working papers, drawings, schedules, tabulations,

22   projections, information or programs stored in a computer (whether or not ever

23   printed out or displayed), and All drafts, alterations, modifications, changes or

24   amendments of any of the foregoing;

25   •   Graphic or aural representations of any kind, Including, without limitation,

26   photographs, microfiche, microfilm, videotapes, recordings, drawings, charts and

27   motion pictures;

28   •   All letters, words, pictures, sounds, or symbols, or combinations thereof stored in or

PLAINTIFFS' FIRST SET OF INTERROGATORIES TO DEFENDANT META

on any electronic, mechanical, magnetic, or optical device Including, but not limited to: (i) computer data storage devices (servers, laptops hard-drives, flash drives, discs, magnetic cards, and the like), (ii) the internet or "the Cloud" (such as e-mail, web posts, social media posts, internet pages, etc.), and (iii) information stored on cell phones.

8.   "Identify" as it pertains to Persons means to describe each Person by name, residence address, residence telephone number, occupation, title, business address, and business telephone number. The term "Identify" as pertains to Documents means to state, to the extent known, the date the Document bears, if any; the title of the Document; the author(s) of the Document; the recipient(s) of the Document and the present location(s) or custodian of the Document.

9.   "Including" and "Includes" are used to provide examples of certain types of information and should not be construed as limiting a request or definition in any way. The terms "Including" and "Includes" shall be construed as if followed by the phrase "but not limited to."

10.   "Llama 1" means the series of language models originally styled "LLaMA" and released by Meta in February 2023, as well as precursor models and variant models.

11.   "Llama 2" means the series of language models of the same name released by Meta in July 2023, as well as precursor models and variant models.

12.   "Llama 3" means the series of language models of the same name intended as a successor to Llama 2 and currently in development at Meta (according to a public statement by Meta in October 2023), as well as precursor models and variant models.

13.   "Meta" means Meta Platforms, Inc., their respective parents, owners, directors, subsidiaries and any company, business entity or person in which any of them possess an ownership interest greater than five percent.

14.   "Meta Language Models" means the series of large language models known as Llama 1, Llama 2, and Llama 3, as well as their precursor models and variant models.

15.   "Person" means any natural person or any business, legal, or governmental entity or association.

PLAINTIFFS' FIRST SET OF INTERROGATORIES TO DEFENDANT META

1 COOLEY LLP
   BOBBY GHAJAR (198719)
2 (bghajar@cooley.com)
   COLETTE GHAZARIAN (322235)
3 (cghazarian@cooley.com)
   1333 2nd Street, Suite 400
4 Santa Monica, California 90401
   Telephone: (310) 883-6400
5
   COOLEY LLP
6 MARK WEINSTEIN (193043)
   (mweinstein@cooley.com)
7 KATHLEEN HARTNETT (314267)
   (khartnett@cooley.com)
8 JUDD LAUTER (290945)
   (jlauter@cooley.com)
9 ELIZABETH L. STAMESHKIN (260865)
   (lstameshkin@cooley.com)
10 3175 Hanover Street
   Palo Alto, CA 94304-1130
11 Telephone: (650) 843-5000

12 CLEARY GOTTLIEB STEEN & HAMILTON LLP
   ANGELA L. DUNNING (212047)
13 (adunning@cgsh.com)
   1841 Page Mill Road, Suite 250
14 Palo Alto, CA 94304
   Telephone: (650) 815-4131
15
   [Full Listing on Signature Page]
16
   Counsel for Defendant Meta Platforms, Inc.
17

18   **UNITED STATES DISTRICT COURT**

19   **NORTHERN DISTRICT OF CALIFORNIA**

20 RICHARD KADREY, et al.,   Case No. 3:23-cv-03417-VC

21   Individual and Representative Plaintiffs,   **DEFENDANT META PLATFORMS, INC.'S**
                                                     FURTHER SUPPLEMENTAL AND AMENDED
22   v.   **RESPONSES AND OBJECTIONS TO**
                                                     PLAINTIFFS' FIRST SET OF
23 META PLATFORMS, INC., a Delaware   **INTERROGATORIES**
   corporation;
24
25   Defendant.   Trial Date: None
                                                     Date Action Filed: July 7, 2023
26

27

28

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

PROPOUNDING PARTY:	**PLAINTIFFS RICHARD KADREY, SARAH SILVERMAN, CHRISTOPHER GOLDEN, TA-NEHISI COATES, JUNOT DÍAZ, ANDREW SEAN GREER, DAVID HENRY HWANG, MATTHEW KLAM, LAURA LIPPMAN, RACHEL LOUISE SNYDER, JACQUELINE WOODSON, LYSA TERKEURST, AND CHRISTOPHER FARNSWORTH**	
RESPONDING PARTY:	**DEFENDANT META PLATFORMS, INC.**	
SET NUMBER: ONE	**ONE**	

Pursuant to Federal Rule of Civil Procedure 33 and Local Rule 33, Defendant Meta Platforms, Inc. ("Meta") responds as follows to Plaintiffs Richard Kadrey, Sarah Silverman, Christopher Golden, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Jacqueline Woodson, Lysa TerKeurst, and Christopher Farnsworth's ("Plaintiffs") First Set of Interrogatories ("Interrogatories").

I.   RESPONSES TO ALL INTERROGATORIES

1.   Meta's responses to these Interrogatories are made to the best of Meta's current employees' present knowledge, information, and belief. Said responses are at all times subject to such additional or different information that discovery or further investigation may disclose and, while based on the present state of Meta's recollection, is subject to such refreshing of recollection, and such additional knowledge of facts, as may result from Meta's further discovery or investigation. Meta reserves the right to make any use of, or to introduce at any hearing and at trial, information and/or documents responsive to these Interrogatories but discovered subsequent to the date of these responses, including, but not limited to, any such information or documents obtained in discovery herein.

2.   To the extent that Meta responds to an Interrogatory by stating that Meta will provide information or documents that Meta deems to embody material that is private, business confidential, proprietary, trade secret, or otherwise protected from disclosure pursuant to Federal Rule of Civil Procedure 26(c) and/or Federal Rule of Evidence 501, Meta will only do so subject to the parties' stipulated protective order governing the unauthorized use or disclosure of such information or documents with a designated "CONFIDENTIAL" or "HIGHLY CONFIDENTIAL

1 - ATTORNEYS' EYES ONLY" or "HIGHLY CONFIDENTIAL – SOURCE CODE" distinction

2 (ECF No. 90, the "Protective Order").

3   **3.**   Meta reserves all objections or other questions as to the competency, relevance,

4 materiality, privilege or admissibility as evidence in any subsequent proceeding in or trial of this

5 or any other action for any purpose whatsoever of Meta's responses herein and any document or

6 thing identified or provided in response to the Interrogatories.

7 **II.   OBJECTIONS TO INSTRUCTIONS AND DEFINITIONS[1]**

8   Whether or not separately set forth in response to each Interrogatory, Meta makes these

9 objections to the following Instructions and Definitions:

10   **1.**   Meta objects to the definition of "Agreement" as overbroad and unduly burdensome

11 to the extent that it encompasses oral contracts, arrangements, or understandings, including those

12 that are informal. Meta further objects to the definition of "Agreement" as vague, ambiguous, and

13 unintelligible as to the term "modifications" to the extent it is intended to mean something distinct

14 from "versions" or "amendments." Meta will construe "Agreement" to mean written contracts,

15 including drafts, versions, amendments, exhibits, and appendices thereof.

16   **2.**   Meta objects to the definition of "Communications" to the extent it is inconsistent

17 with and otherwise seeks to circumvent the custodian and search term limits for electronic

18 communications (including emails and other electronic correspondence, and documents attached

19 thereto), as provided in the Stipulated Protocol regarding Electronic Discovery ("ESI

20 Order"). Meta will produce Documents, including Communications, pursuant to the terms of the

21 ESI Order, and any agreement to produce such Documents is explicitly in view of the terms of the

22 ESI Order. To the extent that Meta responds to a Request, including by agreeing to search for

23 relevant, non-privileged communications in Meta's possession, custody, or control, such response

24 is not a representation that any particular custodian or search term is appropriate. Meta expressly

25 reserves the right to object to any custodians and search terms proposed by Plaintiffs.

26

27

28

___

[1] In connection with these First Supplemental Responses and Objections to Plaintiffs' First Set of Interrogatories, Meta has updated its objections to Plaintiffs' definitions of "Communications," "Llama 3," and "Meta Language Models" to address the entry of the ESI Order and the release of Llama 3.

3. Meta objects to the definitions of "Llama 1," "Llama 2," and "Llama 3" as vague and ambiguous as to the undefined terms "precursor models" and "variant models." Meta further objects to these definitions as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent that it purports to require Meta to produce documents or information concerning large language models ("LLMs") that were not publicly released and/or were not trained on corpuses of text that include any of Plaintiffs' allegedly copyrighted works. For the same reason, Meta objects to these definitions to the extent that they purport to require Meta to produce documents or information concerning LLMs that are not relevant to any party's claims or defenses. For purposes of these responses, Meta construes the term "Llama 1" to refer to the LLM released by Meta as Llama on February 24, 2023, the term "Llama 2" to refer to the LLM released by Meta under that name on July 18, 2023, and the term "Llama 3" to refer to the LLM released by Meta under that name on April 18, 2024, July 23, 2024, and September 25, 2024.

4. Meta objects to the definition of "Meta" as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent that it purports to require Meta to produce documents or information concerning any "owners" regardless of shareholder interest and shareholders with an ownership of in Meta of greater than 5%. Meta will construe "Meta" or "You" to mean Meta Platforms, Inc.

5. Meta objects to the definition of "Meta Language Models" as vague and ambiguous as to the undefined terms "precursor models" and "variant models." Meta further objects to this definition as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent that it purports to require Meta to produce documents concerning LLMs that were not publicly released and/or were not trained on corpuses of text that allegedly include any of Plaintiffs' allegedly copyrighted works. For the same reason, Meta objects to this definition to the extent that it purports to require Meta to produce documents that are not relevant to any party's claims or defenses. Meta will construe "Meta Language Models" to mean the models within the Llama family of LLMs that have been publicly released by Meta, namely, Llama 1, Llama 2, Code Llama, and Llama 3 (as those terms are construed above).

1   **6.**   Meta objects to the definition of "Relevant Period" as vague, ambiguous, and

2   unintelligible, as it is defined circularly to mean "all times relevant to… the Complaint." Meta will

3   construe the Relevant Period to mean January 1, 2022 to the present.

4   **7.**   Meta objects to the definition of "Training Data" as vague, ambiguous, and

5   unintelligible as to the term "other material," which is indefinite and undefined. Meta further

6   objects to the definition of "Training Data" as vague and ambiguous as to the phrase "considered

7   for use," which, read literally, would encompass any dataset considered by any Meta employee,

8   regardless of the seriousness of such consideration and whether or not that consideration was ever

9   acted upon. Meta further objects to this definition to the extent it purports to include datasets (or

10 "considered" datasets) that include content to which Plaintiffs have made no claim of ownership

11 and which are not the subject of any allegations of copyright infringement by Plaintiffs. Meta will

12 construe "Training Data" to mean the "Books3" textual dataset used to train the Meta Language

13 Models (as construed above).

14   **8.**   Meta objects to the definition of "You" and "Your" as overbroad, unduly

15 burdensome, and nonsensical, insofar as it refers to "the specific Defendant(s) producing

16 documents in response to these Requests." There is only one defendant in this case, Meta, and this

17 response is to the Interrogatories, not any document requests. Meta further objects to this definition

18 to the extent it seeks to impose upon Meta an obligation to investigate information or documents

19 outside of its possession, custody, or control. For purposes of these responses, Meta construes the

20 terms "You" and "Your" coextensively with Meta (as construed above).

21   **9.**   Meta objects to Instruction 1 to the extent that it purports to require more of Meta

22 than any obligation imposed by law, and would subject Meta to unreasonable and undue burden

23 and expense. Meta will supplement or amend its responses to these Interrogatories in accordance

24 with Meta's obligations under Rule 26(e).

25   **10.**   Meta objects to Instruction 2, which defines the "Relevant Period" as January 1,

26 2000 to the present. Such definition is overbroad, unduly burdensome, and disproportionate to the

27 needs of the case because it both precedes the existence of Facebook (and therefore Meta) by

28 several years, and the development of the Meta Language Models by decades. For the same reason,

the definition of "Relevant Period," as applied to the Interrogatories, would encompass information

that is irrelevant to the parties' claims and defenses. The Instruction is also inconsistent with the

definition of "Relevant Period" provided on page 3 of the Interrogatories and is therefore vague

and ambiguous. Meta will construe the Relevant Period to mean January 1, 2022 to the present.

11. Meta objects to Instruction 4 (referring to Fed. R. Civ. P. Rule 33(d)) on the ground

that it purports to require more of Meta than any obligation imposed by law, and would subject

Meta to unreasonable and undue burden and expense.

12. Meta objects to Instruction 6 (outlining additional obligations for allegedly

incomplete responses) to the extent that it purports to require Meta to investigate information

outside of its possession, custody, or control.

13. Meta objects to Instruction 8 (outlining additional obligations for any privilege

objection) on the ground that it purports to require more of Meta than any obligation imposed by

law, and would subject Meta to unreasonable and undue burden and expense.

14. Meta objects to Instruction 9 (outlining additional obligations for any work product

objection) on the ground that it purports to require more of Meta than any obligation imposed by

law, and would itself require disclosure of information protected by attorney-client privilege and/or

attorney work product doctrine.

15. Meta objects to Instruction 10 (building in a separate question for each

Interrogatory) on the ground that it purports to require more of Meta than any obligation imposed

by law, seeks disclosure of information protected by attorney-client privilege and/or attorney work

product doctrine, and seeks to circumvent Plaintiffs' interrogatory limit.

16. Meta objects to Instruction 11 (purporting to require responses for "all predecessors,

successors, subsidiaries … divisions and/or affiliates of Meta"), on the ground that it purports to

require more of Meta than any obligation imposed by law, and would subject Meta to unreasonable

and undue burden and expense. Meta further objects to Instruction 11 to the extent that it purports

to require Meta to investigate information outside of its possession, custody, or control. As such

the Instruction if overly broad, as well. Subject to any objections applicable to a particular

1   Interrogatory, Meta will conduct a reasonable, proportionate search for non-privileged, relevant,

2   responsive information within its possession, custody, or control.

3   **17.**   In responding to all Interrogatories, Meta will comply with the requirements of the

4   Federal Rules of Evidence and Federal Rule of Civil Procedure 26.

5   **III.   OBJECTIONS AND RESPONSES TO INDIVIDUAL INTERROGATORIES**

6   **INTERROGATORY NO. 1:**

7   Describe in detail the data You have used to train or otherwise develop the Meta Language

8   Models, Including, for each:

9   a.   How You obtained the data, e.g., by scraping the data, purchasing it from third parties, or

10   by other means;

11   b.   All sources of Data, including any third parties that provided data sets;

12   c.   To the extent the data was derived from publicly available websites, a list of all such

13   websites and, for each, the percentage of the data corpus that is derived from that website;

14   d.   The categories of content included in the data and the extent to which each category is

15   represented in the data corpus (i.e., as a percentage of data used to train the model);

16   e.   All policies and procedures Related to identifying, assessing, vetting and selecting sources

17   of data for the model.

18   **RESPONSE TO INTERROGATORY NO. 1:**

19   Meta incorporates by reference its objections and definitions above, including to the terms

20   "You" and "Meta Language Models." Meta further notes that the capitalized term "Related" is not

21   defined; Meta construes that term coextensively with "concerning."

22   As an initial matter, Meta objects to this Interrogatory because it consists of multiple,

23   separate Interrogatories, each which count toward Plaintiffs' limit under the Federal Rules. For

24   example, the question about what data used to train a model is separate from how it was obtained,

25   and further, subparts (d) and (e) are not subsumed within and necessarily related to the primary

26   question, and purport to require a calculation of percentages of data, and separate identification of

27   "policies" and "procedures" for (1) identifying, (2) assessing, (3) vetting, and (4) selecting data.

28   This Interrogatory consists of *at least* three Interrogatories, and depending on how it is interpreted,

1   many more. In answering the Interrogatory, Meta does not waive this objection.

2   Meta objects to this Interrogatory because, on its face, it does not exclude legal advice or

3   opinions, which are subject to attorney-client privilege and/or attorney work product doctrine, in

4   particular as to subpart (e). Meta will not produce privileged materials or attorney work product.

5   Meta objects to this Interrogatory as vague and ambiguous as to the term "data," which is

6   undefined. Meta will construe "data" to mean Training Data (as construed above).

7   Meta objects to this Interrogatory as vague, ambiguous, and unintelligible as to "percentage

8   of that data corpus that is derived from that website" because "data corpus" is undefined, and Meta

9   is accordingly unable to interpret and respond to subpart (c). Even if "data corpus" were defined,

10   the subject matter of subpart (c) would be overbroad, unduly burdensome, and disproportionate to

11   the needs of the case and seeks information that is not relevant to the parties' claims and defenses.

12   Meta will not respond to subpart (c).

13   Meta objects to the undefined phrase "categories of content, which is vague, ambiguous,

14   and unintelligible.

15   Meta objects to this Interrogatory to the extent that it seeks information that is not within

16   Meta's possession, custody, or control.

17   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

18   Protective Order and the ESI Order, Meta responds as follows: Meta incorporates by reference the

19   identification of datasets used to train Llama 1 that is included in the publicly available paper

20   "LLaMA: Open and Efficient Foundation Language Models." Such datasets were used to train

21   Llama 1. Meta will produce a copy of that paper in its forthcoming production pursuant to Rule

22   33(d).

23   Meta will conduct a reasonable search for additional non-privileged information or, in

24   accordance with Rule 33(d), documents in Meta's possession, custody, or control, sufficient to

25   show any other datasets used to train the Meta Language Models (as construed above), as well as

26   policies and procedures for identifying, assessing, vetting, and selecting sources of data for those

27   models.

28   Discovery is ongoing and Meta will also supplement its response to this Interrogatory to

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

1 identify the sources of such datasets and general categories of data within them, to the extent that

2 such information is within Meta's possession, custody, or control.

3   Discovery is continuing and Meta reserves the right to supplement or amend its response at

4 a later time.

5 **Meta's First Supplemental and Amended Response to Interrogatory No. 1:**

6   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

7 Protective Order, Meta responds as follows.

8   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

9 **Protective Order.**

10   Based on its reasonable investigation, Meta identifies the following datasets as containing

11 material used to train the Llama Models (as construed above), including pretraining and/or

12 finetuning, as well the locations from which Meta believes they were obtained:

Dataset name	Llama 1	Llama 2	Llama 3	URL or Other Location
Stack Exchange	Yes	Yes	Yes	https://archive.org/details/stackexchange
books3	Yes	Yes	Yes	https://the-eye.eu/public/AI/pile_preliminary_components/books3.tar.gz
Project Gutenberg	Yes	Yes	Yes	https://www.gutenberg.org
Arxiv	Yes	Yes	Yes	https://www.arxiv.org
Github	Yes	Yes	Yes	https://www.github.com
C4	Yes	Yes	Yes	https://www.tensorflow.org/datasets/catalog/c4
CCNet	Yes	Yes	No	https://github.com/speedinghzl/CCNet/blob/master/LICENSE
CC-stories	Yes	Yes	No	https://github/com/tensorflow/models/tree/archive/research/lm_commonsense#1-download-data-files
The Stack	Yes	Yes	Yes	https://huggingface.co/datasets/bigcode/the-stack
Wikipedia	Yes	Yes	Yes	https://en.wikipedia.org/wiki/Wikipedia:Database_download
███████████	No	Yes	Yes	███████████
███████████	No	Yes	No	███████████
███████████	No	Yes	Yes	███████████

	No	Yes	Yes
	No	Yes	Yes
	No	Yes	Yes
	No	Yes	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes
	No	No	Yes

1	██████	No	No	Yes	██████
2		No	No	Yes	
3		No	No	Yes	
4					
5					
6					
7		No	No	Yes	
8	Libgen	No	No	Yes	https://libgen.is
9		No	No	Yes	
		No	No	Yes	
10		No	No	Yes	
11		No	No	Yes	
12		No	No	Yes	
13		No	No	Yes	
		No	No	Yes	
14		No	No	Yes	
		No	No	Yes	
15		No	No	Yes	
16		No	No	Yes	
17		No	No	Yes	
18		No	No	Yes	
19		No	No	Yes	
20		No	No	Yes	
21		No	No	Yes	
22		No	No	Yes	
23		No	No	Yes	
24					
25		No	No	Yes	
26		No	No	Yes	
27		No	No	Yes	
28					

███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███
███	No	No	Yes	███

1   The annotations data used for finetuning Llama 2 are identified in Table 6 of the paper titled

2   "Llama 2: Open Foundation and Fine-Tuned Chat Models." Except for the "Meta (Safety &

3   Helpfulness)" data, which was obtained from Meta's vendors, namely, ▮▮▮▮▮▮▮ these annotations

4   datasets were sourced from publicly available sources, such as Github and Hugging Face. Meta has

5   also entered agreements with ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮, to provide annotations.

6   In addition, for Llama 3.1, Meta used publicly available data sourced from ▮▮▮▮▮▮▮▮▮

7   ▮▮▮▮▮▮▮▮ to train the model, as well as a variety of synthetic data.

8   The process for selecting datasets for use in pre-training of the Meta Language Models (as

9   construed above) was informed by what data was available, whether the development team believed

10   that the data would help the model achieve optimal results against industry benchmarks, and PXFN

11   review, i.e., cross-functional review by legal, privacy, and/or policy personnel. Each of the above

12   external datasets was required to undergo PXFN review prior to training of the Meta Language

13   Models (as construed above). Any issues related to intellectual property are regarded as legal in

14   nature. Review and consideration of those issues is therefore the responsibility of Meta's legal

15   team, rather than Meta privacy or policy personnel, and is subject to attorney-client privilege and/or

16   work product doctrine.

17   From the development team's perspective, decisions around which datasets to use for Llama

18   1 were influenced by the development of other large language models, in particular DeepMind's

19   Chinchilla and the corresponding paper "Training Compute-Optimal Large Language Models." At

20   the time, researchers regarded DeepMind's Chinchilla as state of the art, and the team developing

21   the first version of Llama was motivated to reproduce Chinchilla's results on industry benchmarks

22   (e.g., MMLU, BoolQ, PIQA, etc.) using their own model architecture. Using the same or similar

23   dataset diversity allowed the team to better compare the effectiveness of the respective models.

24   Llama 2 was largely trained on the same datasets as Llama 1.

25   With respect to Llama 3, whether a particular dataset was used in the training of the model

26   was driven by a number of considerations, including:

27   • Dataset size – It is understood that LLMs, such as the Meta Language Models (as construed

28   above), require large volumes of text data in order to achieve high performance across

COOLEY LLP
ATTORNEYS AT LAW

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

1   industry benchmarks. In general, the more data the models train on, the better the

2   performance of the model. That is, there is a rough correlation between the number

3   of unique token strings within a dataset and the performance of the models on

4   downstream tasks, such as the ability to answer questions. Accordingly, larger datasets

5   are preferred to smaller datasets.

6   • Dataset diversity – Datasets with greater diversity of subject matter, a variety of lengths and

7   human/computer languages, and different styles of writing or conversation help enable the

8   models to be more flexible and adaptable to different contexts.

9   • Dataset quality – Related to the diversity of the dataset is the extent to which undesirable

10   data (such as repetitive data, factually incorrect data, or harmful or toxic data) can be filtered

11   from the dataset without degrading dataset usefulness.

12 The data mix that will achieve the best results against benchmarks (e.g., MMLU, GSM8K,

13 BoolQ, PIQA, CommonsenseQA, etc.) is difficult to determine in advance. Accordingly, the Meta

14 Language Model (as construed above) development teams performed small scale experiments prior

15 to full scale pre-training to evaluate optimal data mix proportions. Pursuant to Rule 33(d), Meta

16 also refers Plaintiffs to the paper titled "The Llama 3 Herd of Models," published by Meta on July

17 23, 2024, for further information.

18 **INTERROGATORY NO. 3:**

19 Describe in detail the RLHF process for each Meta Language Model. Include in Your

20 response:

21 a.   Examples of types of experts who write questions and answers for use in RLHF;

22 b.   Examples of questions and answers;

23 c.   An explanation of the rating system or method of evaluation for the Meta Language

24 Model's responses;

25 d.   A description of the RLHF You actually undertook in order to correct or remediate

26 any Meta Language Model's propensity to emit protected expression from its Training Data.

27 **RESPONSE TO INTERROGATORY NO. 3:**

28 Meta incorporates by reference its objections and definitions above, including to the terms

1   "Meta Language Model," "You," and "Your."

2   As an initial matter, Meta objects to this Interrogatory because it consists of multiple,

3   separate Interrogatories, each which count toward Plaintiffs' limit under the Federal Rules. For

4   example, the question about the RLHF process is separate from subpart (b), which asks for

5   "examples of questions answers" or subpart (c) which asks for a description of a rating system.

6   Moreover, subpart (d) asks about steps taken to "correct or remediate" *any* language model's

7   "propensity" to emit "protected expression," which is a separate inquiry altogether.   This

8   Interrogatory consists of *at least* three Interrogatories. In answering the Interrogatory, Meta does

9   not waive this objection.

10   Meta objects to this Interrogatory as vague and ambiguous as to "examples of types of

11   experts." Meta will construe this phrase to refer to examples of the general background of

12   individuals who have written questions and answers for use in connection with RLHF for Meta

13   Language Models (as construed above).

14   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

15   to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

16   in particular as to subparts (a)-(c).

17   Meta objects to this Interrogatory to the extent that it seeks information that is not within

18   Meta's possession, custody, or control.

19   Meta objects to the counterfactual presumption and characterization of its language models

20   as having a "propensity to emit protected expression" from training data. Relatedly, "protected

21   expression" appears to call for a legal conclusion and is otherwise vague.

22   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

23   Protective Order and the ESI Order, Meta responds as follows:

24   **This response is designated as Confidential under the Protective Order.**

25   Meta did not undertake an RLHF process for Llama 1.

26   With respect to Llama 2, Meta undertook an RLHF process, initially with the assistance of

27   ███. For a description of that process,

28   Meta incorporates by reference the detailed description included in the publicly available paper

1   "Llama 2: Open Foundation and Fine Tuned Chat Models." Meta will produce a copy of that paper
2   in its forthcoming production pursuant to Rule 33(d).

3   Additionally, Meta will, following additional investigation, identify documents or
4   information sufficient to show (i) representative examples of questions and answers used during
5   the Llama 2 RLHF process, (ii) the method of evaluation for Llama 2's responses during the RLHF
6   process, and (iii) steps undertaken as a part of RLHF of Llama 2, if any, that reduce the likelihood
7   that the model could reproduce verbatim content from any training data, to the extent such
8   documents or information are within Meta's possession, custody, or control and can be located after
9   a reasonable search.

10   Discovery is continuing and Meta reserves the right to supplement or amend its response at
11   a later time.

12   **Meta's First Supplemental Response to Interrogatory No. 3:**

13   Subject to and without waiving the foregoing objections, and pursuant to the terms of the
14   Protective Order, Meta identifies the following documents as containing examples of questions and
15   answers used during the RLHF process: Meta_Kadrey_00014794-00015376,
16   Meta_Kadrey_00016896-00017243, Meta_Kadrey_00017244-00018017,
17   Meta_Kadrey_00019182-00019625, and Meta_Kadrey_00022649-00022996.

18   Meta has produced a copy of "Llama 2: Open Foundation and Fine-Tuned Chat Models"
19   as Meta_Kadrey_00000001-00000077. Pursuant to Rule 33(d), Meta also refers Plaintiffs to the
20   paper titled "The Llama 3 Herd of Models," published by Meta on July 23, 2024, for further
21   information.

22   **Meta's Second Supplemental Response to Interrogatory No. 3:**

23   **This response is designated as Confidential under the Protective Order.**

24   Meta refers to the deposition testimony of Thomas Scialom which discusses the steps
25   undertaken as a part of RLHF of Llama 2 in order to try to reduce the likelihood that the model
26   could reproduce verbatim content from any training data, namely, that the model was finetuned to
27   respond to requests to reproduce certain types of outputs with pre-written responses. T. Scialom
28   12/05/2024 Dep. Tr. 122:11-125:17; H. Touvron 12/03/2024 Dep. Tr. 257:3-259:6. Similar

1   techniques were adopted for more recently released versions of Llama.

2   **INTERROGATORY NO. 4:**

3   Describe in detail the policies and procedures that You follow in order to assess risk, safety,

4   and alignment before You release a new Meta Language Model to the public, whether paid or free.

5   Include in Your response:

6   a.   A description of the risks taken into consideration, including the risks that a Meta

7   Language Model will emit protected expression from its Training Data;

8   b.   A description of the individual(s), type(s) of individual by title and area of expertise,

9   or department(s) responsible for determining whether the Meta Language Model is ready to be

10   released;

11   c.   A list of all instances where a Meta Language Model was released to the public after

12   passing this review;

13   d.   A list of all instances where a Meta Language Model was not released to the public

14   after failing this review;

15   e.   A list of all instances where a Meta Language Model was released to the public

16   despite failing this review, and which individual(s) were responsible for overriding the result of

17   this review.

18   **RESPONSE TO INTERROGATORY NO. 4:**

19   Meta incorporates by reference its objections and definitions above, including to the terms

20   "Meta Language Model," "You," and "Training Data."

21   As an initial matter, Meta objects to this Interrogatory because it consists of multiple,

22   separate Interrogatories, each which count toward Plaintiffs' limit under the Federal Rules. The

23   Interrogatory consists of *at least* four Interrogatories, which count toward Plaintiffs' limit. For

24   example, subparts (c), (d), and (e) are not subsumed within and necessarily related to the primary

25   question. In answering the Interrogatory, Meta does not waive this objection.

26   Meta objects to this Interrogatory because, on its face, it does not exclude legal advice or

27   opinions, which are subject to attorney-client privilege and/or attorney work product doctrine, in

28   particular as the subject matter of the Interrogatory encompasses policies and procedures that are

1 legal in nature. Meta will not respond to subparts (a), (d), and (e).

2   Meta objects to the phrase "risk, safety, and alignment" as undefined and thus vague and

3 ambiguous. Meta also objects to the phrase "this review" as vague, ambiguous, and undefined,

4 and, in particular, to the extent that it implies anything about the type of "review," if any, undertaken

5 by Meta.

6   Meta objects to the counterfactual presumption and characterization of its language models

7 as having a propensity to "emit protected expression" from training data. Relatedly, "protected

8 expression" appears to call for a legal conclusion and is otherwise vague.

9   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

10 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

11 in particular to the extent it seeks information concerning policies and procedures that have no

12 bearing on issues related to U.S. copyright law.

13   Meta objects to this Interrogatory to the extent that it seeks information that is not within

14 Meta's possession, custody, or control.

15   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

16 Protective Order, Meta responds as follows:

17   **This response is designated as Confidential under the Protective Order.**

18   The models known as Llama 1, Llama 2, and Code Llama have been released by Meta to

19 the public.

20   Several different teams, including Meta's in-house legal team, were involved in assessing

21 risk and product safety issues for the Meta Language Models (as construed above).

22   The Llama 1 and Llama 2 development teams were responsible for assessing alignment (as

23 that term is used in the paper titled "Llama 2: Open Foundation and Fine Tuned Chat Models") of

24 their respective models. Aurelian Gonzalez was the manager of the Llama 1 team; Sergey Edunov

25 was the manager of the Llama 2 team.

26   Meta will, after additional investigation, identify individuals at Meta primarily responsible

27 for assessing risk, safety, and alignment of the Meta Language Models (as construed above) prior

28 to release.

1   Discovery is continuing and Meta reserves the right to supplement or amend its response at

2   a later time.

3   **Meta's First Supplemental Response to Interrogatory No. 4:**

4   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

5   Protective Order, Meta identifies the following additional individuals:

6   **This response is designated as Confidential under the Protective Order.**

7   • Mike Clark (Director, Product Management)   Mr. Clark played a supervisory r[e]

8   in connection with addressing risk and safety issues during development of Llama

9   2 and Llama 3.

10   • Ahmad Al-Dahle (VP Gen AI) – Mr. Al-Dahle oversees Meta's Gen AI team,

11   including during development and release of Llama 2 and 3.

12   • Chaya Nayak (Director, Product Management) – Ms. Nayak played a supervisory

13   role in connection with addressing risk and safety issues during development of

14   Llama 2 and Llama 3.

15   • Joelle Pineau (VP, AI Research) – Ms. Pineau oversaw Meta's Fundamental AI

16   Research ("FAIR") team during development and release of Llama 1.

17   **Meta's Second Supplemental Response to Interrogatory No. 4:**

18   Meta considered a number of risks in conjunction with development and release of the Meta

19   Language Models (as construed above), including risks associated with data privacy, product

20   safety, intellectual property, social risks (e.g., toxicity and bias), and risks that the models could be

21   abused by bad actors, among other risks. Meta made significant investments to mitigate these risks

22   prior to release of the Meta Language Models (as construed above), in particular Llama 2 onward,

23   and all such models were determined by cross-functional stakeholders to be suitable for release.

24   *See e.g.*, M. Clark 11/13/2024 30(b)(6) Dep. 11:15-22, 16:6-18:15, 101:3-103:22.

25   **INTERROGATORY NO. 5:**

26   Identify all Agreements between and among:

27   a.   You and anyone associated, affiliated, or having any connection with data used to

28   train the Meta Language Models;

1   b.   You and directors and officers of Defendants or Related Entities with a more than

2   five percent interest held by directors and officers of Defendants.

3   **RESPONSE TO INTERROGATORY NO. 5:**

4   Meta incorporates by reference its objections and definitions above, including to the terms

5   "Meta Language Models," and "You."

6   As an initial matter, Meta objects to this Interrogatory because it consists of two

7   Interrogatories, which count toward Plaintiffs' limit. In answering the Interrogatory, Meta does not

8   waive this objection.

9   Meta objects to this Interrogatory as vague, ambiguous, and unintelligible as to the phrase

10   "any connection with data." Meta is accordingly unable to respond to subpart (a) as written.

11   Meta objects to this Interrogatory as vague, ambiguous, and unintelligible as to

12   "Defendant*s*," as Meta is the only named defendant in this case. Meta will interpret "Defendants"

13   to mean Meta (as construed above).

14   Meta objects to this Interrogatory as vague, ambiguous, and unintelligible as to the term

15   "Related Entities," which is capitalized as if it is defined when it is not. For purposes of this

16   response, Meta will construe "Related Entities" to mean entities owned by or sharing common

17   ownership with Meta.

18   Meta objects to this Interrogatory on the ground that it is overbroad, unduly burdensome,

19   and disproportionate to the needs of the case and seeks information that is not relevant to the parties'

20   claims and defenses, in particular to the extent it seeks information concerning "all" such

21   agreements regardless of their subject matter having "any connection" with "data." Meta will

22   construe subpart (a) to refer to formal, written agreements concerning licensing or acquisition of

23   training data for any Meta Language Model (as construed above). Meta will not respond to subpart

24   (b).

25   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

26   Protective Order and the ESI Order, Meta responds as follows:

27   Meta incorporates by reference its most recent SEC FORM DEF 14A, which identifies

28   individuals and companies with a beneficial interest of more than 5% of Meta's publicly traded

COOLEY LLP
ATTORNEYS AT LAW

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

1   stock. Meta will produce a copy of that document in its forthcoming production pursuant to Rule

2   33(d).

3   Pursuant to Rule 33(d), Meta will also conduct a reasonable search for and produce on a

4   rolling basis documents sufficient to show any executed written agreements concerning licensing

5   or acquisition of training datasets for any Meta Language Models (as construed above).

6   **Meta's First Supplemental Response to Interrogatory No. 5:**

7   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

8   Protective Order, Meta responses as follows:

9   **This response is designated as Confidential under the Protective Order.**

10   None of the datasets used for pre-training the Meta Language Models (as construed above)

11   were the subject of written, signed agreements. However, annotations data used in finetuning, such

12   as data from ███████ were the subject of agreements with those vendors.

13   **Meta's Second Supplemental Response to Interrogatory No. 5:**

14   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

15   **Protective Order.**

16   Pursuant to Rule 33(d), Meta identifies Meta_Kadrey_00154812, tab 5, as providing the list

17   of agreements that Meta has entered into for purposes of obtaining text, speech, and image data

18   that may be used to train Llama 4. Images licensed from Shutterstock, referenced in rows 58 and

19   59 of tab 7 of Meta_Kadrey_00154812, were also used for the multimodal Llama 3.2 model. In

20   addition, Meta has entered into an agreement with Reuters, ████████████████████████

21   ██

22   **INTERROGATORY NO. 6:**

23   In order of corporate seniority, Identify by name and date all Your past and present directors,

24   officers, and board members.

25   **RESPONSE TO INTERROGATORY NO. 6:**

26   Meta incorporates by reference its objections and definitions above, including to the term

27   "Your."

28   Meta objects to the portion of this Interrogatory requiring Meta to list individuals by order

1   of "corporate seniority," as certain individuals may be of equivalent or otherwise indistinguishable

2   seniority.

3   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

4   to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

5   in particular to the extent it seeks information concerning all past and present directors, officers,

6   and board members of Meta, irrespective of the time period relevant to Plaintiffs' allegations, or

7   the role of said individuals in the conduct at issue in Plaintiffs' complaint, which concerns language

8   models that were only recently released.

9   Meta will not respond to this Interrogatory. Meta instead refers Plaintiff to the Investor

10   Relations   (https://investor.fb.com/home/default.aspx)   and   Company   Info

11   (https://about.meta.com/company-info/) pages on its website, which contain information about

12   present Meta leadership.

13   **INTERROGATORY NO. 7:**

14   In order of corporate seniority, Identify by name, job title, and date, all Persons, Including

15   employees from other businesses, contractors, vendors, and other non-employees of Your business,

16   previously and currently responsible for, or having oversight or control over the training,

17   engineering, development, ethics, safety, or alignment of the Meta Language Models, and any

18   iterations, versions, or variations thereof. Include in Your response:

19   a.   For each Person, a description of his or her area of expertise;

20   b.   For each Person, a description of whether such Person was previously or is currently

21   responsible for, or has or had oversight or control over ethics, safety, or alignment related to the

22   Meta Language Models;

23   c.   For each Person, a description of whether such Person was previously or is currently

24   responsible for, or has or had oversight or control over researching, analyzing, reporting on,

25   mitigating, or remediating the propensity of the Meta Language Models to emit protected

26   expression from the Training Data;

27   d.   For each Person responsive to subpart b above, provide a description of whether

28   such Person created Documents or Communications Concerning the ethics or legality of any

1   Training Data gathered or used by You;

2   e. Provide a description of the mitigations and remediations so undertaken by any

3   Persons responsive to subpart c. above;

4   f. For Persons previously employed, Identify whether they left You voluntarily or

5   involuntarily, and whether any disagreement with You about Training Data was a factor in their

6   separation from You.

7   **RESPONSE TO INTERROGATORY NO. 7:**

8   Meta incorporates by reference its objections and definitions above, including to the terms

9   "You," "Communications," and "Meta Language Models."

10   As an initial matter, Meta objects to this Interrogatory because it consists of *at least* four

11   Interrogatories, which count toward Plaintiffs' limit. For example, subparts (d), (e), and (f) are not

12   subsumed within and necessarily related to the primary question. In answering the Interrogatory,

13   Meta does not waive this objection.

14   Meta objects to this Interrogatory because, on its face, it does not exclude legal advice or

15   opinions, which are subject to attorney-client privilege and/or attorney work product doctrine, in

16   particular as the subject matter of the Interrogatory encompasses policies and procedures that are

17   legal in nature. Meta will not respond to subpart (d). Meta also will not respond to subpart (e) to

18   the extent it seeks information subject to attorney-client privilege and/or attorney work product

19   doctrine.

20   Meta objects to this Interrogatory as vague and ambiguous as to the terms "safety" and

21   "ethics," and the phrase "iterations, versions, or variations thereof," all of which are undefined (and

22   as to "any iterations, versions, or variations thereof" as overly broad and unduly burdensome).

23   Meta also objects to the reference to "date" as vague, ambiguous, and unintelligible as to

24   the subject of this Interrogatory.

25   Meta further objects to the portion of this Interrogatory requiring Meta to list individuals by

26   order of "corporate seniority," as certain individuals may no longer be employed by Meta (and thus

27   have no seniority), while others may be of equivalent or otherwise indistinguishable seniority. Meta

28   will identify individuals in alphabetical order of their last names.

1   Meta objects to this Interrogatory as vague and ambiguous as to the phrase "responsible for,

2 or having oversight or control," which, read broadly, would include a large number of individuals

3 regardless of the significance of their role. Meta will construe this Interrogatory to seek information

4 concerning individuals in a managerial role or who may be regarded as lead developers with

5 primary responsibility for the training, engineering, development, or alignment of the Meta

6 Language Models (as construed above).

7   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

8 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

9 in particular to the extent it seeks information concerning issues unrelated to Plaintiffs' direct

10 copyright infringement claim.

11   Meta objects to this Interrogatory to the extent that it seeks information that is not within

12 Meta's possession, custody, or control, in particular as to individuals who are no longer with the

13 company.

14   Meta objects to the counterfactual presumption and characterization of its language models

15 as having a "propensity to emit protected expression" from training data.

16   **Meta's Further Supplemental and Amended Response to Interrogatory No. 7:**

17   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

18 Protective Order and the ESI Order, Meta responds as follows:

19   Meta refers Plaintiffs to Appendix A.1 to the publicly available paper "Llama 2: Open

20 Foundation and Fine-Tuned Chat Models," which identifies the following individuals as having

21 leadership roles in connection with development of Llama 2:

22 ● **Science and Engineering Leadership:**

23   ○ Guillem Cucurull (Research Engineer, former Meta employee) – Mr. Cucurull left

24   Meta voluntarily. Meta is unaware of any information to suggest that his departure

25   from Meta concerned a disagreement with Meta over training data.

26   ○ Naman Goyal (Software Engineer)

27   ○ Louis Martin (Research Scientist, former Meta employee) – Mr. Martin left Meta

28   voluntarily. Meta is unaware of any information to suggest that his departure from

Meta concerned a disagreement with Meta over training data.

- o Thomas Scialom (Research Scientist)
- o Ruan Silva (Software Engineer)
- o Kevin Stone (Research Engineer, former Meta employee) – Mr. Stone left Meta voluntarily. Meta is unaware of any information to suggest that his departure from Meta concerned a disagreement with Meta over training data.
- o Hugo Touvron (Research Scientist)

- **Technical and Management Leadership:**
  - o Sergey Edunov (Director, AI Research)
  - o Angela Fan (Research Scientist)
  - o Melanie Kambadur (Research Engineering Manager)
  - o Sharan Narang (Research Scientist Manager)
  - o Aurelien Rodriguez (Software Engineering Manager, former Meta employee) – Mr. Rodriguez left Meta voluntarily. Meta is unaware of any information to suggest that his departure from Meta concerned a disagreement with Meta over training data.
  - o Robert Stojnic (Engineering Manager, former Meta employee) – Mr. Stojnic left Meta voluntarily. Meta is unaware of any information to suggest that his departure from Meta concerned a disagreement with Meta over training data.

Meta has produced a copy of the Llama 2: Open Foundation and Fine-Tuned Chat Models paper as Meta_Kadrey_00000001-00000077. Excluding Mr. Silva and Mr. Martin, each of the foregoing listed individuals also contributed to development of Llama 3. Pursuant to Rule 33(d), Meta also refers Plaintiffs to the paper titled "The Llama 3 Herd of Models," published by Meta on July 23, 2024, for further information.

Meta also identifies the following individuals as having leadership roles in connection with development of Llama 2 and Llama 3:

- Mike Clark (Director, Product Management)
- Ahmad Al-Dahle (VP Gen AI)
- Chaya Nayak (Director, Product Management)

In addition, Meta identifies the following individuals as having leadership roles in connection with development of Llama 1:

- Joelle Pineau (VP, AI Research)

- Edouard Grave (Research Scientist, former Meta employee) – Mr. Grave was a research scientist at Meta. Mr. Grave was one of the lead developers of Llama 1. Mr. Grave left Meta voluntarily. Meta is unaware of any information to suggest that his departure from Meta concerned a disagreement with Meta over training data.

- Guillaume Lample (Research Scientist, former Meta employee) – Mr. Lample was a research scientist at Meta. Mr. Lample was one of the lead developers of Llama 1. Mr. Lample left Meta voluntarily. Meta is unaware of any information to suggest that his departure from Meta concerned a disagreement with Meta over training data.

- Aurelien Rodriguez (Software Engineering Manager, former Meta employee)

- Hugo Touvron (Research Scientist)

This response is designated as Highly Confidential – Attorney's Eyes Only under the Protective Order.

With respect to steps taken by Meta to reduce the possibility, if any, that Meta Language Models (as construed above) could reproduce verbatim content contained within its training data, Meta responds as follows:

With respect to Llama 2, Meta undertook the following actions to reduce the likelihood that memorization of training data could occur.

- Deduplicated texts within training data where practicable;

- Trained model with only one epoch or less for certain datasets;

- Application of decoding methods to statistically reduce the likelihood of memorization;

- Used annotations and red teaming during fine-tuning to test for memorization;

- Conducted tests on the model to evaluate the extent to which training data was memorized and determined that the memorization rate of strings of 50 words or

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

1   more was approximately 0% in response to ordinary prompts and less than .2% in

2   response to adversarial prompts.

3   With respect to Llama 3, the process was similar. However, use of datasets was typically

4   not limited to one epoch because Meta determined that, for Llama 3, memorization of content of a

5   given dataset did not appreciably increase with multiple epochs. In addition, Meta improved its

6   dataset deduplication process.

7   Meta refers to the deposition testimony of Thomas Scialom which discusses the steps

8   undertaken as a part of RLHF of Llama 2 in order to try to reduce the likelihood that the model

9   could reproduce verbatim content from any training data, namely, that the model was finetuned to

10   respond to requests to reproduce certain types of outputs with pre-written responses. T. Scialom

11   12/05/2024 Dep. Tr. 122:11-125:17; *see also* H. Touvron 12/03/2024 Dep. Tr. 257:3-259:6.

12   Similar techniques were adopted for more recently released versions of Llama.

13   In addition, for Meta AI, Meta has implemented a feature that compares Meta AI output to

14   a database of music lyrics and, if the platform detects that the output contains song lyrics within

15   the database (as measured by a semantic distance algorithm, and a configured matching threshold),

16   the output is removed from display to the user. *See* Meta_Kadrey_00186180.

17   **INTERROGATORY NO. 8:**

18   Identify all software, databases, or services previously and currently used by You for

19   training, maintaining, supervising, managing, analyzing, programming, updating, troubleshooting,

20   diagnosing, testing or modifying the Meta Language Models.

21   **RESPONSE TO INTERROGATORY NO. 8:**

22   Meta incorporates by reference its objections and definitions above, including to the terms

23   "You" and "Meta Language Models."

24   As an initial matter, Meta objects to this Interrogatory because it consists of twenty-seven

25   Interrogatories, which count toward Plaintiffs' limit. Specifically, "software," "databases" and

26   "services" are all broad, distinct subject matters, as are "training," "maintaining," "supervising,"

27   "managing," "analyzing," "programming," "updating," "troubleshooting [and] diagnosing",

28   "testing," and "modifying." In answering the Interrogatory, Meta does not waive this objection.

COOLEY LLP
ATTORNEYS AT LAW

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

1   Meta objects to this Interrogatory as vague, ambiguous, and unintelligible as to

2 "maintaining," "supervising," "managing," "troubleshooting," "diagnosing," and "modifying" in

3 the context of Meta's Llama models.

4   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

5 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

6 in particular to the extent it seeks information concerning issues unrelated to Plaintiffs' theory of

7 copyright infringement.

8   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

9 Protective Order and the ESI Order, Meta responds as follows:

10   **This response should be treated as Confidential under the Protective Order.**

11   The primary software tool used for development of Llama 1 and 2 was PyTorch.

12 **INTERROGATORY NO. 10:**

13   Identify all Persons or Related Entities from whom You licensed, purchased or otherwise

14 obtained Training Data for the Meta Language Models. For each such provider of Training Data,

15 include the name of such person(s), date, amount paid, approximate description and size of data

16 You obtained, and Identify the particular models such Training Data was used for.

17 **RESPONSE TO INTERROGATORY NO. 10:**

18   Meta incorporates by reference its objections and definitions above, including to the terms

19 "You," "Training Data," and "Meta Language Models."

20   As an initial matter, Meta objects to this Interrogatory because it consists of *at least* three

21 Interrogatories, which count toward Plaintiffs' limit. For example, identifying a "Person" is

22 separate from the date of an alleged transaction, how much was paid, or specifics about data. In

23 answering the Interrogatory, Meta does not waive this objection.

24   Meta objects to this Interrogatory as vague, ambiguous, and unintelligible as to the term

25 "Related Entities" which is capitalized as if it is defined when it is not. Meta will construe "Related

26 Entities" to mean entities owned by or sharing common ownership with Meta.

27   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

28 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

1   in particular to the extent it seeks information concerning issues unrelated to Plaintiffs' theory of

2   copyright infringement.

3   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

4   Protective Order and the ESI Order, Meta responds that it will identify Persons from whom Meta

5   has "licensed, purchased, or otherwise obtained" Training Data (as construed above) for the Meta

6   Language Models (as construed above), if any, to the extent such information is within Meta's

7   possession, custody, or control, after additional investigation has been completed. Discovery is

8   continuing and Meta reserves the right to supplement or amend its response at a later time.

9   **Meta's First Supplemental Response to Interrogatory No. 10:**

10   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

11 Protective Order, Meta responds as follows:

12   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

13 **Protective Order.**

14   Meta obtained datasets used to pre-train the Meta Language Models (as construed above)

15 from the sources identified in Meta's supplemental response to Interrogatory No. 1. Except for the

16 annotations data acquired from Meta's vendors, none of the datasets was the subject of any payment

17 or fees.

18   **Meta's Second Supplemental Response to Interrogatory No. 10:**

19   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

20 **Protective Order.**

21   Pursuant to Rule 33(d), Meta identifies Meta_Kadrey_00154812, tab 5, as providing the list

22 of agreements that Meta has entered into for purposes of obtaining text, speech, and image data

23 that may be used to train Llama 4. Images licensed from Shutterstock, referenced in rows 58 and

24 59 of tab 7 of Meta_Kadrey_00154812, were also used for the multimodal Llama 3.2 model. In

25 addition, Meta has entered into an agreement with Reuters, ████████████████████

26 ██

27 **INTERROGATORY NO. 11:**

28   Identify by name all individuals or entities who possess or have possessed stock or

1 ownership interests in You greater than five percent.

2 **RESPONSE TO INTERROGATORY NO. 11:**

3   Meta incorporates by reference its objections and definitions above, including to the term

4 "You." Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

5 to the needs of the case, is unbounded in time and scope, and seeks information that is not relevant

6 to the parties' claims and defenses. A list of all individuals or entities that have at any point in time

7 possessed stock or ownership in Meta greater than five percent is irrelevant to Plaintiffs' theory of

8 copyright infringement.

9   Meta incorporates by reference its most recent SEC FORM DEF 14A, which identifies

10 individuals and companies with a beneficial interest of more than 5% of Meta's publicly traded

11 stock. Meta will produce a copy of that document in its forthcoming production pursuant to Rule

12 33(d).   Meta   also   refers   Plaintiff   to   the   Investor   Relations

13 (https://investor.fb.com/home/default.aspx) and Company Info (https://about.meta.com/company-

14 info/) pages on its website, which contain information about present Meta leadership.

15 **INTERROGATORY NO. 12:**

16   Identify by name, department, and job description, all individuals who have directly

17 participated in the planning, conception, design, programming, testing, or operation of the Meta

18 Language Models during the Relevant Period (including organization charts if applicable),

19 Including All Persons with responsibility who have directly participated in the choice or selection

20 of Training Data for the Meta Language Models.

21 **RESPONSE TO INTERROGATORY NO. 12:**

22   Meta incorporates by reference its objections and definitions above, including to the terms

23 "Relevant Period," "Training Data," and "Meta Language Models."

24   As an initial matter, Meta objects to this Interrogatory because it consists of several

25 Interrogatories, each which count toward Plaintiffs' limit.

26   Meta objects to this Interrogatory as vague and ambiguous as to the phrase "directly

27 participated in," which, read literally, may include a large number of individuals regardless of the

28 significance of their role. Meta also objects to this Interrogatory as vague and ambiguous as to the

1 term "operation," which could mean any number of things, including any use of a model. Meta

2 will construe this Interrogatory to seek information concerning individuals in a managerial role or

3 who may be regarded as lead developers with primary responsibility for the planning, conception,

4 design, programming, testing, or selection of Training Data (as construed above) used for Meta

5 Language Models (as construed above).

6   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

7 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

8 in particular to the extent it seeks information concerning issues unrelated to Plaintiffs' direct

9 copyright infringement claim.

10   Meta objects to this Interrogatory to the extent that it seeks information that is not within

11 Meta's possession, custody, or control, in particular as to individuals who are no longer with the

12 company and any "organization charts" of relevant employees.

13   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

14 Protective Order and the ESI Order, Meta responds as follows:

15   Meta incorporates by reference Appendix A.1 to the publicly available paper "Llama 2:

16 Open Foundation and Fine-Tuned Chat Models," which identifies the following individuals as

17 having leadership roles in connection with development of Llama 2:

18   ·   **Science and Engineering Leadership:** Guillem Cucurull (Research Engineer),

19 Naman Goyal (Software Engineer), Louis Martin (Research Scientist), Thomas Scialom (Research

20 Scientist), Ruan Silva (Software Engineer), Kevin Stone (Research Engineer), Hugo Touvron (AI

21 Research Scientist).

22   ·   **Technical and Management Leadership:** Sergey Edunov (Director, AI Research),

23 Angela Fan (Research Scientist), Melanie Kambadur (Research Engineering Manager), Sharan

24 Narang (Research Scientist Manager), Aurelien Rodriguez (Software Engineering Manager),

25 Robert Stojnic (Engineering Manager).

26   Meta will produce a copy of that paper in its forthcoming production pursuant to Rule 33(d).

27   **This response is designated as Confidential under the Protective Order.**

28   In addition, Meta identifies the Edouard Grave (Research Scientist, former employee),

1   Guillaume Lample (Research Scientist, former employee), Hugo Touvron (AI Research Scientist),

2   and Aurelien Rodriguez (Software Engineering Manager) as lead developers with primary

3   responsibility for the planning, conception, design, programming, testing, or operation of Llama 1.

4   **Meta's First Supplemental Response to Interrogatory No. 12:**

5   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

6   **Protective Order.**

7   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

8   Protective Order, Meta incorporates herein its Supplemental and Amended Response to

9   Interrogatory No. 7.

10   Meta has produced a copy of "Llama 2: Open Foundation and Fine-Tuned Chat Models" as

11   Bates No. Meta_Kadrey_00000001- Meta_Kadrey_00000077. Pursuant to Rule 33(d), Meta also

12   refers Plaintiffs to the paper titled "The Llama 3 Herd of Models," published by Meta on July 23,

13   2024, for further information.

14   **INTERROGATORY NO. 13:**

15   Describe in detail the policies and procedures that You follow or have followed in assessing

16   whether to permit the use of a dataset as Training Data for a Meta Language Model, and:

17   a.   list all datasets that have been so reviewed;

18   b.   list datasets that have been permitted to be used as Training Data for any Meta

19   Language Model, and the date so designated;

20   c.   list datasets that have at any time been forbidden from being used as Training Data

21   for any Meta Language Model;

22   d.   for each forbidden dataset, list the reason that it was forbidden and the date so

23   forbidden;

24   e.   if a forbidden dataset was later designated safe to use, list the reason its designation

25   was revised and the date so revised.

26   **RESPONSE TO INTERROGATORY NO. 13:**

27   Meta incorporates by reference its objections and definitions above, including to the terms

28   "You," "Training Data," and "Meta Language Models."

1   As an initial matter, Meta objects to this Interrogatory because it consists of at least three

2   Interrogatories, which count toward Plaintiffs' limit. Specifically, questions about so-called

3   policies or procedures are separate from a list of datasets reviewed, reasons why a dataset was

4   "forbidden" (a term used by the Interrogatory), or reasons why a dataset was once "forbidden" but

5   later not. In answering the Interrogatory, Meta does not waive this objection.

6   Meta objects to this Interrogatory because, on its face, it does not exclude legal advice or

7   opinions, which are subject to attorney-client privilege and/or attorney work product doctrine, in

8   particular as the subject matter of the Interrogatory encompasses policies and procedures that are

9   legal in nature. Meta will not provide information concerning policies and procedures that

10   constitute legal advice, nor will Meta respond to subparts (a), (b), to the extent that it concerns

11   whether and when a given dataset received approval for use, and (c)-(e).

12   Meta objects to this Interrogatory as vague and ambiguous as to the term "forbidden

13   dataset," which is undefined. Meta further objects to this term to the extent it implies that Meta has

14   prohibited use of any particular dataset for training a language model. Meta will construe

15   "forbidden dataset" to mean datasets referred to in subpart (c).

16   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

17   to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

18   in particular to the extent it seeks information concerning issues unrelated to Plaintiffs' direct

19   copyright infringement claim.

20   Meta objects to this Interrogatory as duplicative and cumulative of Interrogatory No. 1, and

21   therefore refers to its response to Interrogatory 1.

22   Meta objects to this Interrogatory because it exceeds Plaintiffs' limit of 25 Interrogatories

23   under Rule 33(a)(1).

24   **Meta's First Supplemental Response to Interrogatory No. 13:**

25   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

26   Protective Order, the ESI Order, and Rule 33(d), Meta identifies Bates Nos.

27   Meta_Kadrey_00089197,   Meta_Kadrey_00093496,   Meta_Kadrey_00065402,   and

28   Meta_Kadrey_00154774 as identifying datasets that Meta has considered for use as text training

1 data for the Meta Language Models (as construed above).

2 **INTERROGATORY NO. 14:**

3   Identify by name all individuals or entities who have applied for access to Llama 1. Include

4 in Your response:

5   a.   The names all individuals or entities who obtained access to Llama 1;

6   b.   The names all individuals or entities who you denied access to Llama 1, including

7 the names of all individuals or entities who have had their access to Llama 1 revoked or otherwise

8 limited by You;

9   c.   For each individual or entity whom you denied access to Llama 1, a description of

10 the reason(s) why you denied access to that particular individual or entity.

11 **RESPONSE TO INTERROGATORY NO. 14:**

12   Meta incorporates by reference its objections and definitions above, including to the terms

13 "You," "Your," and "Llama 1."

14   As an initial matter, Meta objects to this Interrogatory because it consists of at least three

15 Interrogatories, which count toward Plaintiffs' limit. Specifically, an inquiry regarding entities

16 who applied for access is separate from a question about who obtained or was denied access, and

17 the reasons why (on a person by person basis).

18   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

19 to the needs of the case as it would require Meta to compile a list of every individual who sought

20 to use Llama v1 and when and whether they were given access (or not), and then determine the

21 reasoning (if there was one) why access was or was not granted. There is no relevance or

22 proportionality of this request to Plaintiffs' direct copyright infringement claim relating to training.

23   Meta objects to this Interrogatory because, on its face, it does not exclude legal advice or

24 opinions, which are subject to attorney-client privilege and/or attorney work product doctrine, in

25 particular as the subject matter of the Interrogatory encompasses determinations that are legal in

26 nature.

27   Finally, Meta objects to this Interrogatory because it exceeds Plaintiffs' limit of 25

28 Interrogatories under Rule 33(a)(1).

COOLEY LLP
ATTORNEYS AT LAW

33

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' FIRST SET OF ROGS
3:23-CV-03417-VC

1   **Meta's First Supplemental Response to Interrogatory No. 14:**

2   Pursuant to Rule 33(d), Meta identifies the Bates Nos. Meta_Kadrey_00065897,

3   Meta_Kadrey_00065898,   Meta_Kadrey_00065899,   Meta_Kadrey_00066056,

4   Meta_Kadrey_00066065,   Meta_Kadrey_00187042,   Meta_Kadrey_00187043,

5   Meta_Kadrey_00187044, Meta_Kadrey_00187045. These documents identify individuals who

6   received, and who were denied, access to Llama 1.   Meta also identifies Bates Nos.

7   Meta_Kadrey_00185688, Meta_Kadrey_00185680, which describe the criteria Meta used to grant

8   or deny access to Llama 1.

9   **INTERROGATORY NO. 15:**

10   Explain in detail the role of the following individuals in any aspect of the design, training,

11   development, testing, marketing, release or support of each of the Meta Language Models:

12   a.   Hugo Touvron

13   b.   Aurélien Rodriguez

14   c.   Tim Dettmers

15   d.   Luke Zettlemoyer

16   e.   Shawn Presser

17   f.   Stella Biderman

18   g.   Leo Gao

19   h.   Yann LeCun

20   i.   John Carmack

21   j.   Andrew Bosworth

22   k.   Chris Cox

23   l.   Jennifer Newstead

24   m.   Ahmad Al-Dahle

25   n.   Marc Zuckerberg

26   o.   Marc Andreessen

27   **RESPONSE TO INTERROGATORY NO. 15:**

28   Meta incorporates by reference its objections and definitions above, including to the term

1   "Meta Language Models."

2   As an initial matter, Meta objects to this Interrogatory because it consists of at least fifteen

3   Interrogatories, which count toward Plaintiffs' limit. Specifically, each individual's role in the

4   broad subjects of "design, training, development, testing, marketing, release or support" are

5   separate subjects, and are not necessarily related to one another.

6   Meta objects to this Interrogatory as vague and ambiguous as to the term "support," which

7   is undefined and effectively meaningless. Meta also objects to this Interrogatory as vague and

8   ambiguous as to the phrase "involved in any aspect of," which is undefined and overbroad. Meta

9   will construe this Interrogatory to seek information concerning the identified individuals' job

10   responsibilities in connection with the "design, training, development, testing, marketing, and

11   release" of the Meta Language Models (as construed above).

12   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

13   to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

14   in particular to the extent it seeks information concerning issues unrelated to Plaintiffs' theory of

15   copyright infringement.

16   Finally, Meta objects to this Interrogatory because it exceeds Plaintiffs' limit of 25

17   Interrogatories under Rule 33(a)(1).

18   **Meta's First Supplemental Response to Interrogatory No. 15:**

19   Based on a reasonable investigation, Meta describes the roles of the following individuals

20   in the development of the Meta Language Models as follows:

21   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

22   **Protective Order.**

23   • <u>Hugo Touvron</u> – Mr. Touvron helped develop the architecture of Llama beginning with

24   Llama 1, and oversaw experimentation and ablations for Llama 1. For Llama 2, Mr. Touvron also

25   co-supervised RLHF. For Llama 3, Mr. Touvron was responsible for exploratory pretraining. Mr.

26   Touvron also worked on mixture of experts ("MOE") model training, which was not used for Llama

27   3, but may be used for Llama 4. For Llama 4, Mr. Touvron is also working on developing computer

28   keyboard and mouse use capabilities.

1   • Aurélien Rodriguez – Mr. Rodriguez was the manager for the Llama 1 development

2   team. In that role, his responsibilities included organizing the team's work and general approach,

3   supervising training, and keeping the team on task. For Llama 2, Mr. Rodriguez was a part of the

4   technical and Management leadership wherein he oversaw pretraining and was in charge of

5   supervised finetuning. For Llama 3, Mr. Rodriguez worked for a time on pretraining, but turned

6   his focus to research for development of Llama 4.

7   • Tim Dettmers – Mr. Dettmers was not involved in the development of the Meta Language

8   Models.

9   • Luke Zettlemoyer – Mr. Zettlemoyer was not involved in the development of the Meta

10   Language Models.

11   • Shawn Presser – Mr. Presser was not involved in the development of the Meta Language

12   Models.

13   • Stella Biderman – Ms. Biderman was not involved in the development of the Meta

14   Language Models.

15   • Leo Gao – Mr. Gao was not involved in the development of the Meta Language Models.

16   • Yann LeCun – Mr. LeCun had limited involvement in the development of the Meta

17   Language Models. He was one of the founding members of the FAIR group at Meta, helped

18   assemble a team of AI researchers, and provided input to further Meta's AI research efforts. Mr.

19   LeCun also provided input into Meta's decision to release the Meta Language Models under an

20   open-source license.

21   • John Carmack – Mr. Carmack was not involved in the development of the Meta Language

22   Models.

23   • Andrew Bosworth – Mr. Bosworth, as Chief Technology Officer, oversaw Meta's FAIR

24   group as part of Meta's Reality Labs Division until FAIR was restructured into Mr. Cox's

25   organization in early 2024 (see below). Mr. Bosworth provided high-level oversight over AI

26   development and research at Meta and input on Meta's open-source licensing policy for the Meta

27   Language Models and corporate strategy around investment in and development of AI models and

28   solutions.

- <u>Chris Cox</u> – Mr. Cox, as Chief Product Officer, oversees Meta's Generative AI (GenAI) group and (as of early 2024) Meta's FAIR group, and therefore provides high-level oversight over development of the Meta Language Models, including the use of Meta Language Models for Meta products and services such as Meta AI. He also provides input on Meta's open-source licensing policy for the Meta Language Models and Meta's corporate strategy around investment in and development of AI models and solutions.

- <u>Jennifer Newstead</u> – Ms. Newstead, in conjunction with and supported by others on Meta's legal team, provided legal guidance during the development of the Meta Language Models concerning a variety of legal issues, including open source licensing, data privacy, intellectual property, regulatory issues, and other legal risks.

- <u>Ahmad Al-Dahle</u> – Mr. Al-Dahle oversees Meta's GenAI group. He was not involved in the development of Llama 1, but as head of GenAI oversaw the development of Llama 2 and subsequent Meta Language Models. Pursuant to Rule 33(d), Meta refers to the transcript of the deposition of Mr. Al-Dahle on October 3, 2024.

- <u>Mark Zuckerberg</u> – Mr. Zuckerberg, as CEO of Meta, provides high level management and oversight of the company and corporate strategy, including in connection with its AI technologies, products, and strategies. Mr. Zuckerberg has also been involved in promoting the Meta Language Models to strategic partners and the broader public, and has publicly supported and advocated for an open-source licensing approach for the Meta Language Models.

- <u>Marc Andreessen</u> – Mr. Andreessen is a member of Meta's Board of Directors and was present during some of the Board meetings in which Meta's plans for the Meta Language Models were discussed. He was not involved in the development of the Meta Language Models.

1   Dated: December 13, 2024   COOLEY LLP

2

                                  By: */s/ Judd Lauter*

3   Bobby Ghajar

                                   Mark Weinstein

4   Kathleen Hartnett

                                   Phillip Morton

5   Judd Lauter

                                   Elizabeth L. Stameshkin

6   Matthew Brigham

                                   Colette Ghazarian

7   Juan Pablo Gonzalez

                                   Cole A. Poppell

8

                                   LEX LUMINA PLLC

9   Mark A. Lemley

10   CLEARY GOTTLIEB STEEN &
                                   HAMILTON LLP

11   Angela L. Dunning

12   Attorneys for Defendant
                                   META PLATFORMS, INC.

13   *Full Counsel List*

14   COOLEY LLP

15   PHILLIP MORTON *(pro hac vice)*
     (pmorton@cooley.com)

16   COLE A. POPPELL *(pro hac vice)*
     (cpoppell@cooley.com)

17   1299 Pennsylvania Avenue, NW, Suite 700
     Washington, DC 20004-2400

18   Telephone: (202) 842-7800

19   COOLEY LLP
     MATTHEW BRIGHAM (191428)

20   (mbrigham@cooley.com)
     JUAN PABLO GONZALEZ (334470)

21   (jgonzalez@cooley.com)
     3175 Hanover Street

22   Palo Alto, CA 94304-1130
     Telephone: (650) 843-5000

23   LEX LUMINA PLLC

24   MARK A. LEMLEY (155830)
     (mlemley@lex-lumina.com)

25   745 Fifth Avenue, Suite 500
     New York, NY 10151

26   Telephone: (646) 898-2055

27

28

1

PROOF OF SERVICE

2   I am a citizen of the United States and a resident of the State of California. I am employed

3   in Los Angeles County, State of California, in the office of a member of the bar of this Court, at

4   whose direction the service was made. I am over the age of eighteen years, and not a party to the

5   within action. My business address is Cooley LLP, 355 South Grand Avenue, Suite 900, Los

6   Angeles, CA 90071. On the date set forth below I served the documents described below in the

7   manner described below:

8   • **DEFENDANT META PLATFORMS, INC.'S FURTHER SUPPLEMENTAL AND AMENDED RESPONSES AND OBJECTIONS TO PLAINTIFFS' FIRST SET OF INTERROGATORIES**

9

10 ☒ (BY ELECTRONIC MAIL) I am personally and readily familiar with the business practice of Cooley LLP for the preparation and processing of documents in portable document format (PDF) for e-mailing, and I caused said documents to be prepared in PDF and then served by electronic mail to the parties listed below.

11

12 on the following part(ies) in this action:

13

     Executed on December 13, 2024, at Los Angeles, California.

14

15   */s/Jerry Gonzalez*
                                                _____

16   Jerry Gonzalez

17

18

19

20

21

22

23

24

25

26

27

28

1	**SERVICE LIST**
2	Joseph R. Saveri
	Joshua I. Schiller
	Maxwell Vaughn Pritt
3	Cadio Zirpoli
	Christopher K.L. Young
	BOIES SCHILLER FLEXNER
4	Holden Benon
	44 Montgomery Street, 41st Floor
	Louis Andrew Kessler
	San Francisco, CA 94104
5	Aaron Cera
	Email: jischiller@bsfllp.com
6	Margaux Poueymirou
	JOSEPH SAVERI LAW FIRM, LLP
	mpritt@bsfllp.com
7	601 California Street, Suite 1000
	San Francisco, CA 94108
	David Boies (*pro hac vice*)
8	Email: jsaveri@saverilawfirm.com
	BOIES SCHILLER FLEXNER
	czirpoli@saverilawfirm.com
	333 Main Street
9	cyoung@saverilawfirm.com
	Armonk, NY 10504
	hbenon@saverilawfirm.com
	Email: dboies@bsfllp.com
10	lkessler@saverilawfirm.com
	acera@saverilawfirm.com
	Jesse Panuccio (*pro hac vice*)
11	mpoueymirou@saverilawfirm.com
	BOIES SCHILLER FLEXNER
	1401 New York Ave. NW
12		Washington, DC 20005
	Matthew Butterick
	Email: jpanuccio@bsfllp.com
13	MATTHEW BUTTERICK,
	ATTORNEY AT LAW
	Attorneys for Individual and Representative
14	1920 Hillhurst Avenue, #406
	Plaintiffs and the Proposed Class
	Los Angeles, CA 90027
15	Email: mb@buttericklaw.com
	Brian O'Mara
	DiCELLO LEVITT LLP
16		4747 Executive Drive, Suite 240
	Attorneys for Individual and Representative
	San Diego, CA 92121
17	*Plaintiffs and the Proposed Class*
	Email: BrianO@dicellolevitt.com
18	Bryan L. Clobes (*pro hac vice*)
	Alexander J. Sweatman (*pro hac vice*)
	Amy Keller (*pro hac vice*)
19	Mohammed Rathur
	James A. Ulwick (*pro hac vice*)
	Nada Djordjevic (*pro hac vice*)
20	CAFFERTY CLOBES MERIWETHER
	& SPRENGEL LLP
	DiCELLO LEVITT LLP
	10 North Dearborn St., Sixth Floor
21	135 South LaSalle Street, Suite 3210
	Chicago, IL 60603
	Chicago, IL 60602
22	Email: bclobes@caffertyclobes.com
	Email: akeller@dicellolevitt.com
	asweatman@caffertyclobes.com
	julwick@dicellolevitt.com
23	mrathur@caffertyclobes.com
	ndjordjevic@dicellolevitt.com
24	*Attorneys for Individual and Representative*
	David A. Straite (*pro hac vice*)
	Plaintiffs and the Proposed Class
	DiCELLO LEVITT LLP
25		485 Lexington Avenue, Suite 1001
		New York, NY 10017
26		Email: dstraite@dicellolevitt.com
27		*Attorneys for Plaintiff Lysa TerKeurst*
28		

1	Elizabeth J. Cabraser
2	Daniel M. Hutchinson
	Reilly T. Stoler
3	LIEFF CABRASER HEIMANN &
	BERNSTEIN, LLP
4	275 Battery Street, 29th Floor

Elizabeth J. Cabraser
Daniel M. Hutchinson
Reilly T. Stoler
LIEFF CABRASER HEIMANN &
BERNSTEIN, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Telephone: (415) 956-1000
Email: ecabraser@lchb.com
        dhutchinson@lchb.com
        rstoler@lchb.com

Rachel Geman
LIEFF CABRASER HEIMANN &
BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, New York 10013-1413
Telephone: (212) 355-9500
Email: rgeman@lchb.com

*Attorneys for Plaintiff Christopher
Farnsworth and Representative Plaintiffs
and the Proposed Class*

Nancy Evelyn Wolff
COWAN DEBAETS ABRAHAMS &
SHEPPARD LLP
9454 Wilshire Boulevard, Suite 901
Beverly Hills, CA 90212
Telephone: (310) 340-6334
Fax: (310) 492-4394
Email: NWolff@cdas.com

Scott J. Sholder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS &
SHEPPARD LLP
60 Broad Street, 30th Floor
New York, New York 10004
Telephone: (212) 974-7474
Email: ssholder@cdas.com
        ccole@cdas.com

*Attorneys for Plaintiff Christopher
Farnsworth and Representative Plaintiffs
and the Proposed Class*

1   COOLEY LLP
     BOBBY GHAJAR (198719)
2   (bghajar@cooley.com)
     COLETTE GHAZARIAN (322235)
3   (cghazarian@cooley.com)
     1333 2nd Street, Suite 400
4   Santa Monica, California 90401
     Telephone: (310) 883-6400
5

6   COOLEY LLP
     MARK WEINSTEIN (193043)
7   (mweinstein@cooley.com)
     KATHLEEN HARTNETT (314267)
8   (khartnett@cooley.com)
     JUDD LAUTER (290945)
9   (jlauter@cooley.com)
     ELIZABETH L. STAMESHKIN (260865)
10   (lstameshkin@cooley.com)
     3175 Hanover Street
11   Palo Alto, CA 94304-1130
     Telephone: (650) 843-5000

12   CLEARY GOTTLIEB STEEN & HAMILTON LLP
     ANGELA L. DUNNING (212047)
13   (adunning@cgsh.com)
     1841 Page Mill Road, Suite 250
14   Palo Alto, CA 94304
     Telephone: (650) 815-4131
15
     [Full Listing on Signature Page]
16
     Counsel for Defendant Meta Platforms, Inc.
17

18   **UNITED STATES DISTRICT COURT**

19   **NORTHERN DISTRICT OF CALIFORNIA**

20   RICHARD KADREY, et al.,   Case No. 3:23-cv-03417-VC

21   Individual and Representative Plaintiffs,   **DEFENDANT META PLATFORMS, INC.'S**
                                          FURTHER SUPPLEMENTAL AND AMENDED
22   v.   **RESPONSES AND OBJECTIONS TO**
                                          PLAINTIFFS' SECOND SET OF
23   META PLATFORMS, INC., a Delaware   **INTERROGATORIES**
     corporation;
24

25   Defendant.   Trial Date:
                                      Date Action Filed: July 7, 2023
26

27

28

PROPOUNDING PARTY:	**PLAINTIFFS RICHARD KADREY, SARAH SILVERMAN, CHRISTOPHER GOLDEN, TA-NEHISI COATES, JUNOT DÍAZ, ANDREW SEAN GREER, DAVID HENRY HWANG, MATTHEW KLAM, LAURA LIPPMAN, RACHEL LOUISE SNYDER, JACQUELINE WOODSON, LYSA TERKEURST, AND CHRISTOPHER FARNSWORTH**
RESPONDING PARTY:	**DEFENDANT META PLATFORMS, INC.**
SET NUMBER: ONE SECOND	

Pursuant to Federal Rule of Civil Procedure 33 and Local Rule 33, Defendant Meta Platforms, Inc. ("Meta") responds as follows to Plaintiffs Richard Kadrey, Sarah Silverman, Christopher Golden, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Jacqueline Woodson, Lysa TerKeurst, and Christopher Farnsworth's ("Plaintiffs") Second Set of Interrogatories ("Interrogatories").

I.   RESPONSES TO ALL INTERROGATORIES

1.   Meta's responses to these Interrogatories are made to the best of Meta's current employees' present knowledge, information, and belief. Said responses are at all times subject to such additional or different information that discovery or further investigation may disclose and, while based on the present state of Meta's recollection, is subject to such refreshing of recollection, and such additional knowledge of facts, as may result from Meta's further discovery or investigation. Meta reserves the right to make any use of, or to introduce at any hearing and at trial, information and/or documents responsive to these Interrogatories but discovered subsequent to the date of these responses, including, but not limited to, any such information or documents obtained in discovery herein.

2.   To the extent that Meta responds to an Interrogatory by stating that Meta will provide information or documents that Meta deems to embody material that is private, business confidential, proprietary, trade secret, or otherwise protected from disclosure pursuant to Federal Rule of Civil Procedure 26(c) and/or Federal Rule of Evidence 501, Meta will only do so subject to the parties' stipulated protective order governing the unauthorized use or disclosure of such information or documents with a designated "CONFIDENTIAL" or "HIGHLY CONFIDENTIAL

1   - ATTORNEYS' EYES ONLY" or "HIGHLY CONFIDENTIAL – SOURCE CODE" distinction

2   (ECF No. 90, the "Protective Order").

3   **3.**   Meta reserves all objections or other questions as to the competency, relevance,

4   materiality, privilege or admissibility as evidence in any subsequent proceeding in or trial of this

5   or any other action for any purpose whatsoever of Meta's responses herein and any document or

6   thing identified or provided in response to the Interrogatories.

7   **II.   OBJECTIONS TO INSTRUCTIONS AND DEFINITIONS**

8   Whether or not separately set forth in response to each Interrogatory, Meta makes these

9   objections to the following Instructions and Definitions:

10   **1.**   Meta objects to all defined terms to the extent that they are not utilized in Plaintiffs

11   Second Set of Interrogatories.

12   **2.**   Meta objects to the definition of "Agreements" as overbroad and unduly

13   burdensome to the extent that it encompasses oral contracts, arrangements, or understandings,

14   including those that are informal. Meta further objects to the definition of "Agreements" as vague,

15   ambiguous, and unintelligible as to the term "modifications" to the extent it is intended to mean

16   something distinct from "versions" or "amendments." Meta will construe "Agreements" to mean

17   written contracts, including drafts, versions, amendments, exhibits, and appendices thereof.

18   **3.**   Meta objects to the definition of "Communications" to the extent it is inconsistent

19   with and otherwise seeks to circumvent the custodian and search term limits for electronic

20   communications (including emails and other electronic correspondence, and documents attached

21   thereto), as provided in the Stipulated Protocol regarding Electronic Discovery ("ESI

22   Order"). Meta will produce Documents, including Communications, pursuant to the terms of the

23   ESI Order, and any agreement to produce such Documents is explicitly in view of the terms of the

24   ESI Order. To the extent that Meta responds to a Request, including by agreeing to search for

25   relevant, non-privileged communications in Meta's possession, custody, or control, such response

26   is not a representation that any particular custodian or search term is appropriate. Meta expressly

27   reserves the right to object to any custodians and search terms proposed by Plaintiffs.

28

COOLEY LLP
ATTORNEYS AT LAW

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' SECOND SET OF ROGS
3:23-CV-03417-VC

1   **4.**   Meta objects to the definition of "Complaint" which refers to an outdated complaint

2 that has since been replaced by the Corrected Second Consolidated Amended Complaint (ECF No.

3 133). Meta will construe "Complaint" to refer to the Corrected Second Consolidated Amended

4 Complaint.

5   **5.**   Meta objects to the definitions of "Llama 1," "Llama 2," and "Llama 3" as vague

6 and ambiguous as to the undefined terms "precursor models" and "variant models." Meta further

7 objects to these definitions as overbroad, unduly burdensome, and disproportionate to the needs of

8 the case to the extent that it purports to require Meta to produce documents or information

9 concerning large language models ("LLMs") that were not publicly released and/or were not trained

10 on corpuses of text that include any of Plaintiffs' allegedly copyrighted works. For the same reason,

11 Meta objects to these definitions to the extent that they purport to require Meta to produce

12 documents or information concerning LLMs that are not relevant to any party's claims or

13 defenses. For purposes of these responses, Meta construes the term "Llama 1" to refer to the LLM

14 released by Meta as Llama on February 24, 2023, the term "Llama 2" to refer to the LLM released

15 by Meta under that name on July 18, 2023, and the term "Llama 3" to refer to the LLM released by

16 Meta under that name on April 18, 2024, July 23, 2024, and September 25, 2024.

17   **6.**   Meta objects to the definition of "Meta" as overbroad, unduly burdensome, and

18 disproportionate to the needs of the case to the extent that it purports to require Meta to produce

19 documents or information concerning any "owners" regardless of shareholder interest and

20 shareholders with an ownership of in Meta of greater than 5%. Meta will construe "Meta" or "You"

21 to mean Meta Platforms, Inc.

22   **7.**   Meta objects to the definition of "Meta Language Models" as vague and ambiguous

23 as to the undefined terms "precursor models" and "variant models." Meta further objects to this

24 definition as overbroad, unduly burdensome, and disproportionate to the needs of the case to the

25 extent that it purports to require Meta to produce documents concerning LLMs that were not

26 publicly released and/or were not trained on corpuses of text that allegedly include any of Plaintiffs'

27 allegedly copyrighted works. For the same reason, Meta objects to this definition to the extent that

28 it purports to require Meta to produce documents that are not relevant to any party's claims or

1   defenses. Meta will construe "Meta Language Models" to mean the models within the Llama

2   family of LLMs that have been publicly released by Meta, namely, Llama 1, Llama 2, Code Llama,

3   and Llama 3 (as those terms are construed above).

4   **8.**   Meta objects to the definition of "Relevant Period" as vague, ambiguous, and

5   unintelligible, as it is defined circularly to mean "all times relevant to… the Complaint." Meta will

6   construe the Relevant Period to mean January 1, 2022 to the present.

7   **9.**   Meta objects to the definition of "Training Data" as vague, ambiguous, and

8   unintelligible as to the term "other material," which is indefinite and undefined. Meta further

9   objects to the definition of "Training Data" as vague and ambiguous as to the phrase "considered

10   for use," which, read literally, would encompass any dataset considered by any Meta employee,

11   regardless of the seriousness of such consideration and whether or not that consideration was ever

12   acted upon. Meta further objects to this definition to the extent it purports to include datasets (or

13   "considered" datasets) that include content to which Plaintiffs have made no claim of ownership

14   and which are not the subject of any allegations of copyright infringement by Plaintiffs. Meta will

15   construe "Training Data" to mean the "Books3" textual dataset used to train the Meta Language

16   Models (as construed above).

17   **10.**   Meta objects to the definition of "You" and "Your" as overbroad, unduly

18   burdensome, and nonsensical, insofar as it refers to "the specific Defendant(s) producing

19   documents in response to these Requests." There is only one defendant in this case, Meta, and this

20   response is to the Interrogatories, not any document requests. Meta further objects to this definition

21   to the extent it seeks to impose upon Meta an obligation to investigate information or documents

22   outside of its possession, custody, or control. For purposes of these responses, Meta construes the

23   terms "You" and "Your" coextensively with Meta (as construed above).

24   **11.**   Meta objects to Instruction 1 to the extent that it purports to require more of Meta

25   than any obligation imposed by law, and would subject Meta to unreasonable and undue burden

26   and expense. Meta will supplement or amend its responses to these Interrogatories in accordance

27   with Meta's obligations under Rule 26(e).

28

COOLEY LLP
ATTORNEYS AT LAW

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' SECOND SET OF ROGS
3:23-CV-03417-VC

1   **12.**   Meta objects to Instruction 2, which defines the "Relevant Period" as January 1,

2   2000 to the present. Such definition is overbroad, unduly burdensome, and disproportionate to the

3   needs of the case because it both precedes the existence of Facebook (and therefore Meta) by

4   several years, and the development of the Meta Language Models by decades. For the same reason,

5   the definition of "Relevant Period," as applied to the Interrogatories, would encompass information

6   that is irrelevant to the parties' claims and defenses. The Instruction is also inconsistent with the

7   definition of "Relevant Period" provided on page 3 of the Interrogatories and is therefore vague

8   and ambiguous. Meta will construe the Relevant Period to mean January 1, 2022 to the present.

9   **13.**   Meta objects to Instruction 4 (referring to Fed. R. Civ. P. Rule 33(d)) on the ground

10   that it purports to require more of Meta than any obligation imposed by law, and would subject

11   Meta to unreasonable and undue burden and expense.

12   **14.**   Meta objects to Instruction 6 (outlining additional obligations for allegedly

13   incomplete responses) to the extent that it purports to require Meta to investigate information

14   outside of its possession, custody, or control.

15   **15.**   Meta objects to Instruction 8 (outlining additional obligations for any privilege

16   objection) on the ground that it purports to require more of Meta than any obligation imposed by

17   law, and would subject Meta to unreasonable and undue burden and expense.

18   **16.**   Meta objects to Instruction 9 (outlining additional obligations for any work product

19   objection) on the ground that it purports to require more of Meta than any obligation imposed by

20   law, and would itself require disclosure of information protected by attorney-client privilege and/or

21   attorney work product doctrine.

22   **17.**   Meta objects to Instruction 10 (building in a separate question for each

23   Interrogatory) on the ground that it purports to require more of Meta than any obligation imposed

24   by law, seeks disclosure of information protected by attorney-client privilege and/or attorney work

25   product doctrine, and seeks to circumvent Plaintiffs' interrogatory limit.

26   **18.**   Meta objects to Instruction 11 (purporting to require responses for "all predecessors,

27   successors, subsidiaries … divisions and/or affiliates of Meta"), on the ground that it purports to

28   require more of Meta than any obligation imposed by law, and would subject Meta to unreasonable

and undue burden and expense. Meta further objects to Instruction 11 to the extent that it purports to require Meta to investigate information outside of its possession, custody, or control. As such the Instruction if overly broad, as well. Subject to any objections applicable to a particular Interrogatory, Meta will conduct a reasonable, proportionate search for non-privileged, relevant, responsive information within its possession, custody, or control.

19.   In responding to all Interrogatories, Meta will comply with the requirements of the Federal Rules of Evidence and Federal Rule of Civil Procedure 26.

III.   OBJECTIONS AND RESPONSES TO INDIVIDUAL INTERROGATORIES

INTERROGATORY NO. 16:

State all facts on which you base Your contention that Your conduct constitutes fair use (17 U.S.C. § 107).

FIRST AMENDED AND SUPPLEMENTAL RESPONSE TO INTERROGATORY NO. 16:

Meta incorporates by reference its objections and definitions above.

Meta objects to this Interrogatory as vague and ambiguous as to the phrase "Your conduct," which is undefined and could refer to any conduct. Meta will construe this Interrogatory to seek information concerning Meta's claim of fair use in connection with the conduct alleged in the Complaint (as construed above).

Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent it seeks information that Meta does not intend to rely on to support a claim of fair use and calls for a lengthy narrative with regard to twelve different plaintiffs and more than forty works.

Meta objects to this Interrogatory to the extent it prematurely calls for expert testimony or identification of facts yet to be disclosed by Plaintiffs, and to the extent that it requires Meta to respond to legal arguments or theories not yet disclosed by Plaintiffs.

Finally, Meta objects to this Interrogatory because it exceeds Plaintiffs' limit of 25 Interrogatories under Rule 33(a)(1).

Subject to and without waiving the foregoing objections, and pursuant to the terms of the Protective Order, Meta responds as follows:

COOLEY LLP
ATTORNEYS AT LAW

6

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' SECOND SET OF ROGS
3:23-CV-03417-VC

The alleged use of Plaintiffs' asserted works, to the extent it is shown to have occurred, is highly transformative in nature, because it adds something new, with a further purpose or different character, altering those works with considerable new expression, meaning, or message. To the extent Plaintiffs' works were used to train the Meta Language Models, the purpose was transformative both in terms of <u>purpose</u> and <u>expression</u>. With respect to <u>purpose</u>, Plaintiffs' works were allegedly used as data to train the models. In other words, the works were allegedly a part of a corpus of text (specifically, terabytes of text from a variety of sources), from which the models built complex statistical representations of language derived from the patterns, structures, and relationships between words within the corpus. This enables the models to predict the next word in a sequence, and thereby provide useful responses to any manner of input prompts. Such use of textual material serves a fundamentally different purpose from the original texts on which the Meta Language Models were trained. *See e.g.*, Meta_Kadrey_00000001-00000077, Meta_Kadrey_00000078-00000104, Meta_Kadrey_00000224-00000248, Meta_Kadrey_00093669- Meta_Kadrey_00093760.

The text corpus used to train the Meta Language Models includes a large amount of textual materials, and to the extent Plaintiffs' works were used to train those models, they would constitute a tiny fraction of the textual training dataset (both individually and collectively). The purpose of the models, and the use of text datasets, is to create new, original textual output, not to reproduce the content of the datasets with which it was trained. This is demonstrated by, among other things, Meta's efforts to minimize the models' ability to memorize and/or output training data verbatim (*see e.g.*, Meta_Kadrey_00000277) and the wide variety of uses that have been made of the models. *See e.g.*, Meta_Kadrey_00092978-00093308, Meta_Kadrey_00062157. Indeed, all Plaintiffs have admitted that they are not aware of any output from any Meta Language Model that replicates any protected expression in their at-issue books. *See e.g.*, T. Coates 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24 (admitting, subject to objections, that Plaintiff is "personally unaware of any text generated by any of Meta's Llama models that infringes [Plaintiff's] ASSERTED WORKS."); J. Díaz 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24 (same); C. Golden 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to

1   RFA No. 24 (same); A. Greer 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA

2   No. 24 (same); D. Hwang 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No.

3   24 (same); R. Kadrey 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24

4   (same); M. Klam 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24

5   (same); L. Lippman 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24

6   (same); S. Silverman 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24

7   (same); R. Snyder 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24

8   (same); J. Woodson 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 24

9   (same); C. Golden 9/17/2024 Dep. 260:6–261:13 (Mr. Golden testifying that the present lawsuit is

10   not about what comes out of Meta's large language models); A. Greer 9/24/2024 79:21–80:3 (Mr.

11   Greer testifying that his claims concern the use of his works to train large language models); D.

12   Hwang 9/16/2024 Dep. 252:23–253:6 (Mr. Hwang testifying that the operative complaint does not

13   allege that the Meta Language models create any output that is similar to any of his books or plays);

14   L. TerKeurst 9/23/2024 Dep. 226:20–25, 229:5–12 (Ms. TerKeurst testifying that she was not

15   aware of any text generated by the Meta Language Models that was substantially similar that of her

16   asserted works); J. Woodson Dep. 328:23–329:4 (Ms. Woodson testifying that she is not aware of

17   any output from any Meta LLM in which any of her characters appeared.); S. Silverman 10/10/2024

18   Dep. 42:5–8 (Q: "[D]oes it matter if Meta's models never output any language from your book?" A:

19   "It doesn't matter at all."); *id.* 156:25–157:2 ("It doesn't matter what he does with it or what output

20   comes from it. It's not right."); *id.* 321:9–11 ("It's not about the output. If the output might not be

21   this book, but without this book it wouldn't have the out-book [SIC] . . . ."); T. Coates 11/21/2024

22   Dep. 52:22–53:11 (Mr. Coates testifying that he has not personally created or witnessed someone

23   create output from Meta's AI model that replicates or regurgitates portions of his books); M. Klam

24   Rough Drft. Dep. 38:11–16, 234:5–11 (Mr. Klam testifying that he is unaware of any instance in

25   which any Meta AI tool has output verbatim text from any of his books and cannot identify any

26   language he has written in any of his books that was reproduced in the output of Meta's AI model);

27   R. Snyder Rough Drft. Dep. 36:2–4 (Ms. Snyder testifying that she has never seen any output

28   generated by a Meta AI model that copies any language of hers); *cf.* D. Hwang 9/16/2024 Dep.

363:6–15 (Mr. Hwang testifying that he did not believe when he filed suit that the Meta Language Models could create works that were substantially similar to his works); L. Lippman 9/17/2024 Dep. 311:16–312:1 (Ms. Lippman testifying that she is aware that the current version of the Complaint does not allege that Meta's generative AI tools create any output that is substantially similar to her books); A. Greer 9/24/2024 Dep. 28:17–20 (same); J. Díaz 11/20/2024 Dep. 216:4–13 (same)

The transformative nature also extends to <u>expression</u>. The pre-training process involves a number of steps that fundamentally transform the input dataset text in order to facilitate training of the large language model ("LLM"), which Meta will further describe in further detail in connection with expert discovery. At a high-level, the pre-training process includes a "tokenization" step in which the data in training datasets is broken down and encoded into a series of values known as "tokens" which are used to create numerical vector representations that the LLM training algorithms can understand. (*See e.g.*, Meta_Kadrey_00000078-00000104, Section 2.1 ("Tokenizer"); Meta_Kadrey_00000001-00000077, at 6 ("Tokenizer").) The input data is then used in a complex series of LLM training algorithms that adjust the large number of numerical values (known as parameters which include weights) in the LLM, that define the connections and relationships between the nodes in the LLM. By adjusting these weights, the LLM can "learn" and better predict correct outputs based on input data. These numerical parameters enable the LLM, after the training process, to generate better output data in response to input prompts. The process of training of an LLM represents a complete transformation into a form that is entirely unrecognizable from the original training data.

The transformativeness of Meta's use also extends to the post-training and fine-tuning processes used with the Meta Language Models. Plaintiffs have not alleged that their works were used as data in any post-training and fine-tuning processes for any Meta Language Model, but nevertheless, post-training and fine-tuning processes similarly involve a highly transformative use of the data used in those processes, both in terms of purpose and expression. The data used in post-training and fine-tuning is used to fine tune model parameters to improve the performance, quality, and behavior of the models and their responses. For example, in the post-training stage, "the model

1   is tuned to follow instructions, align with human preferences, and improve specific capabilities (for

2   example, coding and reasoning)." The Llama 3 Herd of Models, p.1. In the post-training and fine-

3   tuning processes (which will be discussed in more detail in expert discovery), post-training data is

4   likewise used in a series of complex LLM training algorithms that further tailor the model

5   parameters to improve the quality of responses and the ability of the model to perform various

6   tasks. Post-training and fine-tuning processes also generally involve an amount of data that

7   constitutes a fraction of the amount of data used in pre-training the model. Additionally, and for

8   the same reasons, the transformativeness of Meta's use also extends to research and evaluations

9   (including ablation experiments) to assess the behavior and performance of the Meta Language

10   Models. Plaintiffs likewise have not alleged that their works were used as data for research or

11   evaluation of Meta Language Model, but nevertheless, the data used in these processes serves the

12   transformative purpose of studying and improving model behavior, and furthering the research and

13   development of the Meta Language Models.

14   As a further indication of the transformative nature, and the fact that the training process

15   does not simply make a copy of the input dataset, the training process is so computationally

16   complex that it requires an enormous amount of computing power. *See e.g.*,

17   Meta_Kadrey_00000001-00000077, at 4 ("When training a 65B-parameter model, our code

18   processes around 380 tokens/sec/GPU on 2048 A100 GPU with 80GB of RAM. This means that

19   training over our dataset containing 1.4T tokens takes approximately 21 days.").

20   Aside from its highly transformative nature, Meta's alleged use also has substantial non-

21   commercial, nonprofit, and educational (including research) purposes. Meta is investing billions

22   of dollars in research and development of state-of-the-art LLM technology that it is then making

23   available to the public. In particular, the training of the Meta Language Models resulted in Meta's

24   release of Llama 1, Llama 2, CodeLlama, and Llama 3, 3.1, and 3.2 to the open source

25   community. These LLMs were provided openly to the public, pursuant to an open license that

26   permits developers, researchers, and institutions (with the exception of licensees with more than

27   700 million monthly active users) to use and modify the Llama models free of charge. *See e.g.*,

28   Meta_Kadrey_00000160-00000162,   Meta_Kadrey_00093275-00093284,

1   Meta_Kadrey_00093658-00093760. The open release of Llama has resulted in the Llama models

2   being downloaded hundreds of millions of times by researchers and developers from around the

3   world and has catalyzed development of new and improved AI tools and technologies. *See e.g.*,

4   Meta_Kadrey_00092978-00093308, Meta_Kadrey_00062157. The open and public release of

5   Llama has allowed the public to access highly capable LLM technologies that would otherwise be

6   available only to large organizations and/or at considerable expense. More broadly, Meta's

7   investment and open release is contributing to the U.S. economy, the emergence of a new and

8   important industry, and the U.S.'s global leadership of that industry over geopolitical

9   competitors. In that regard, Meta has agreed to permit members of the Five Eyes intelligence

10   alliance, namely, the United States, Canada, UK, Australia, and New Zealand, to use the Meta

11   Language Models. *See* e.g., Meta_Kadrey_00213585.

12   Plaintiffs' allegedly infringed works were published prior to Meta's alleged use of those

13   works. And Meta's use of large volumes of texts was necessary to extract data regarding, e.g.,

14   word frequencies, grammar, and syntax from those works to generate new content. This

15   information constitutes either facts and ideas, which are not protectable by copyright, or is

16   otherwise unrelated to the purpose of copyright protection. Furthermore, given that training the

17   Meta Language Models requires terabytes of text, that greater volumes of text tends to improve

18   model performance on objective benchmarks measuring reasoning and knowledge of facts, and the

19   formats in which the text is available, it was reasonable for Meta to utilize copies of entire works

20   (as opposed to portions thereof), in particular given Meta's efforts to develop the models in a

21   manner that minimizes the likelihood that training data can be reproduced as model output. For

22   example, books data comprised only approximately 4.5% of tokens used to train Llama 1 and only

23   4.4% of tokens used to train Llama 2.

24   The ordinary market for Plaintiffs' works is the market for people to purchase and read the

25   books and, possibly, to create derivative works from those books, such as for audio books and film

26   and television adaptations. The Meta Language Models and their outputs do not serve as a market

27   substitute for the Plaintiffs' asserted works, do not compete with those works, and do not harm the

28   value of Plaintiffs' asserted works. Plaintiffs have produced no evidence to the contrary, such as

1   evidence of lost sales or other financial harm. To the contrary, plaintiffs have admitted that they

2   are not aware of any such harm. *See e.g.*, T. Coates 9/6/2024 Resps. & Objs. to Meta's 2nd Set of

3   RFAs, Resp. to RFA No. 15 (admitting, subject to objections, that Plaintiff is unaware of lost sales

4   due to alleged infringement); J. Díaz 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to

5   RFA No. 15 (same); C. Golden 8/28/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to

6   RFA No. 15 (same); A. Greer 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA

7   No. 15 (same); D. Hwang 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No.

8   15 (same); R. Kadrey 8/28/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15

9   (same); M. Klam 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15 (same);

10   L. Lippman 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15 (same); S.

11   Silverman 8/28/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15 (same); R.

12   Snyder 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15 (same); J.

13   Woodson 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15 (same); L.

14   TerKeurst 9/12/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 15 (similar);

15   A. Greer 9/6/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 16 (admitting,

16   subject to objections, that Plaintiff's book sales have not decreased due to the alleged use of

17   Plaintiff's Asserted Works to train large language models); L. TerKeurst 9/6/2024 Resps. & Objs.

18   to Meta's 2nd Set of RFAs, Resp. to RFA No. 16 (similar); T. Coates 9/19/2024 Resps. & Objs. to

19   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (admitting, subject to objections, that Plaintiff is

20   unaware of lost licensing opportunities due to alleged infringement); J. Díaz 9/19/2024 Resps. &

21   Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); C. Golden 9/19/2024 Resps. &

22   Objs. to Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); A. Greer 9/19/2024 Resps. & Objs.

23   to Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); D. Hwang 9/19/2024 Resps. & Objs. to

24   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); R. Kadrey 9/19/2024 Resps. & Objs. to

25   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); M. Klam 9/19/2024 Resps. & Objs. to

26   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); L. Lippman 9/19/2024 Resps. & Objs. to

27   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); S. Silverman 9/19/2024 Resps. & Objs. to

28   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); R. Snyder 9/19/2024 Resps. & Objs. to

1   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); J. Woodson 9/19/2024 Resps. & Objs. to

2   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); L. TerKeurst 9/12/2024 Resps. & Objs. to

3   Meta's 2nd Set of RFAs, Resp. to RFA No. 18 (same); T. Coates 7/22/2024 Resps. & Objs. to

4   Meta's 2nd Set of RFAs, Resp. to RFA Nos. 22 and 23 (admitting, subject to objections, that

5   Plaintiff is unaware of (1) persons reading text generated by Meta's Llama models as a substitute

6   for Plaintiff's Asserted Works as described in RFA 22, and (2) documentary evidence of persons

7   reading text generated by Meta's Llama models as substitute for Plaintiff's Asserted Works as

8   described in RFA 23); J. Diaz 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA

9   Nos. 22 and 23 (same); C. Golden 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to

10   RFA Nos. 22 and 23 (same); A. Greer 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs, Resp.

11   to RFA Nos. 22 and 23 (same); D. Hwang 7/22/2024 Resps. & Objs. to Meta's 2nd Set of RFAs,

12   Resp. to RFA Nos. 22 and 23 (same); R. Kadrey 7/22/2024 Resps. & Objs. to Meta's 2nd Set of

13   RFAs, Resp. to RFA Nos. 22 and 23 (same); M. Klam 7/22/2024 Resps. & Objs. to Meta's 2nd Set

14   of RFAs, Resp. to RFA Nos. 22 and 23 (same); L. Lippman 7/22/2024 Resps. & Objs. to Meta's

15   2nd Set of RFAs, Resp. to RFA Nos. 22 and 23 (same); S. Silverman 7/22/2024 Resps. & Objs. to

16   Meta's 2nd Set of RFAs, Resp. to RFA Nos. 22 and 23 (same); R. Snyder 7/22/2024 Resps. & Objs.

17   to Meta's 2nd Set of RFAs, Resp. to RFA Nos. 22 and 23 (same); J. Woodson 7/22/2024 Resps. &

18   Objs. to Meta's 2nd Set of RFAs, Resp. to RFA Nos. 22 and 23 (same); L. TerKeurst 8/21/2024

19   Resps. & Objs. to Meta's 2nd Set of RFAs, Resp. to RFA Nos. 22 and 23 (similar); D. Hwang

20   9/16/2024 Dep. 254:7–13 (Mr. Hwang testifying that he did not know whether he had experienced

21   any financial loss as a result of the alleged training of the Meta Language Models on his asserted

22   works); D. Hwang 9/16/2024 Dep. 290:10–17, 291:22–292:4 (Mr. Hwang testifying that he was

23   not aware of any lost sales of his asserted works due to the alleged infringement in the complaint);

24   R. Kadrey 9/25/2024 Dep. 223:23–224:1 (Mr. Kadrey testifying that he was unaware of any injury

25   other than purported harm of not receiving compensation from Meta); R. Kadrey 9/25/2024 Dep.

26   222:6–9 (Mr. Kadrey testifying that he was not aware of any lost sales of his asserted works due to

27   the alleged infringement in the complaint); L. Lippman 9/17/2024 Dep. 339:3–10 (Ms. Lippman

28   testifying that she is unaware of any financial harm that she has suffered as a result of conduct by

Meta alleged in the Complaint); A. Greer 9/24/2024 Dep. 120:21–121:10 (Mr. Greer testifying that he is unaware of any lost book sales or lost licensing opportunities as a result of Meta's alleged infringement of his works, or whether it is possible that book sales have actually increased due to his participation in this lawsuit); S. Silverman 10/10/2024 Dep. 204:12–205:7, 296:10–297:2 (Ms. Silverman testifying that she is unaware of any evidence to suggest someone did not buy her book because they could generate a summary on a Meta AI tool, evidence of lost sales or lost licensing opportunities due to Meta's conduct, or any instance in which another person did not seek to license her book because of Meta's actions); T. Coates 11/21/2024 Dep. 123:19–125:25 (Mr. Coates testifying that he is unaware of any lost sales or lost licensing opportunities due to Meta's LLMs); J. Díaz 11/20/2024 Dep. 337:22–339:6, 340:21–350:16 (Mr. Díaz testifying that he is unaware of any decrease in sales of his books or any monetary harm suffered due to Meta's use of his books to train its Llama models) ; R. Snyder 12/11/2024 Rough Drft. Dep. 249:22–250:1, 250:14 251:1–253:5, 259:19–23 (Ms. Snyder testifying that she is unaware of any evidence of lost sales or lost licensing opportunities as a result of Meta's conduct alleged in the Complaint); J. Woodson 9/30/2024 Dep. 383:14–20, 389:5–20 (Ms. Woodson testifying that she is unaware of any lost licensing opportunities, loss of income, or lost sales as a result of Meta's conduct alleged in the Complaint); M. Klam 12/10/2024 Rough Drft. Dep. 325:21–326:3, 328:6-10, 330:13–16 (Mr. Klam testifying that he is unaware of any evidence that he lost sales of his asserted works or lost licensing opportunities as a result of Meta's conduct alleged in the Complaint); C. Farnsworth 12/4/2024 Rough Drft. Dep. 138:20–139:8 (Mr. Farnsworth testifying that he could not point to any specific lost opportunity he suffered due to Meta's release of its Llama model).

Moreover, there was no market for licensing Plaintiffs' literary books as training data for LLMs at the time Meta first Llama 1 or at the time Plaintiffs' filed their lawsuit, and there is no such market today. Plaintiffs have admitted as much in discovery. Moreover, testimony from Meta's witnesses, including, by way of example, from Sy Choudhury, and Alex Boesenberg, indicate that there has not been, and is not a market for licensing books for AI training. *See e.g.*, S. Choudhury 12/5/2024 Dep. 20:6-22:21, 65:10-21; A. Boesenberg 11/18/2024 Dep. 381:14-22. Moreover, any particular book has no independent value as training data and/or is

interchangeable with countless other books. In any case, such a market would be for a transformative use. It would also be impractical, if not impossible, for companies developing LLMs to attempt to negotiate licenses with each individual book rights holder for various reasons, including but not limited to the amount of time and cost necessary to do so would have precluded development of the models in the first instance, in particular given the time and costs relative to the de minimis value of individual works to the development of the models. In support, Meta intends to rely on Plaintiffs' discovery responses and testimony, expert reports and testimony, percipient witness testimony, as well as documents produced in this litigation

INTERROGATORY NO. 17:

If You or any of Your employees and/or agents intend to assert the advice of counsel defense, state any and all facts upon which You or any of your employees and/or agents intend to rely on for that contention.

SECOND AMENDED RESPONSE TO INTERROGATORY NO. 17:

Meta incorporates by reference its objections and definitions above.

Meta objects to this Interrogatory as vague and ambiguous as to the reference to "Your employees and/or agents" with respect to any defense in this case, as no Meta employees or agents are parties to this case.

Meta objects to this Interrogatory because it exceeds Plaintiffs' limit of 25 Interrogatories under Rule 33(a)(1).

Subject to and without waiving the foregoing objections, Meta responds as follows: Meta does not intend to assert the advice of counsel defense in this case.

1   Dated: December 13, 2024   COOLEY LLP

2   By: /s/ Judd Lauter

3   Bobby Ghajar
                                                     Mark Weinstein
4   Kathleen Hartnett
                                                     Phillip Morton
5   Judd Lauter
                                                     Elizabeth L. Stameshkin
6   Matthew Brigham
                                                     Colette Ghazarian
7   Juan Pablo Gonzalez
                                                     Cole A. Poppell

8   LEX LUMINA PLLC
9   Mark A. Lemley

10   CLEARY GOTTLIEB STEEN &
                                                     HAMILTON LLP
11   Angela L. Dunning

12   Attorneys for Defendant
                                                     META PLATFORMS, INC.
13 *Full Counsel List*

14 COOLEY LLP
    PHILLIP MORTON *(pro hac vice)*
15 (pmorton@cooley.com)
    COLE A. POPPELL *(pro hac vice)*
16 (cpoppell@cooley.com)
    1299 Pennsylvania Avenue, NW, Suite 700
17 Washington, DC 20004-2400
    Telephone: (202) 842-7800
18
    COOLEY LLP
19 MATTHEW BRIGHAM (191428)
    (mbrigham@cooley.com)
20 JUAN PABLO GONZALEZ (334470)
    (jgonzalez@cooley.com)
21 3175 Hanover Street
    Palo Alto, CA 94304-1130
22 Telephone: (650) 843-5000

23 LEX LUMINA PLLC
    MARK A. LEMLEY (155830)
24 (mlemley@lex-lumina.com)
    745 Fifth Avenue, Suite 500
25 New York, NY 10151
    Telephone: (646) 898-2055

26

27

28

COOLEY LLP
ATTORNEYS AT LAW

META'S FURTHER SUPP. & AMENDED RESP
& OBJ'S TO PLTFS' SECOND SET OF ROGS
3:23-CV-03417-VC

1

PROOF OF SERVICE

2   I am a citizen of the United States and a resident of the State of California. I am employed

3   in Los Angeles County, State of California, in the office of a member of the bar of this Court, at

4   whose direction the service was made. I am over the age of eighteen years, and not a party to the

5   within action. My business address is Cooley LLP, 355 South Grand Avenue, Suite 900, Los

6   Angeles, CA 90071. On the date set forth below I served the documents described below in the

7   manner described below:

8   • **DEFENDANT META PLATFORMS, INC.'S FURTHER SUPPLEMENTAL AND AMENDED
       RESPONSES AND OBJECTIONS TO PLAINTIFFS' SECOND SET OF
9   INTERROGATORIES**

10 ☒ (BY ELECTRONIC MAIL) I am personally and readily familiar with the business
       practice of Cooley LLP for the preparation and processing of documents in
11   portable document format (PDF) for e-mailing, and I caused said documents to be
       prepared in PDF and then served by electronic mail to the parties listed below.

12
    on the following part(ies) in this action:
13

14   Executed on December 13, 2024, at Los Angeles, California.

15

16   */s/Jerry Gonzalez*

17   Jerry Gonzalez

18

19

20

21

22

23

24

25

26

27

28

COOLEY LLP
ATTORNEYS AT LAW

PROOF OF SERVICE
3:23-CV-03417-VC

1

SERVICE LIST

2

Joseph R. Saveri
Cadio Zirpoli
Christopher K.L. Young
Holden Benon
Louis Andrew Kessler
Aaron Cera
Margaux Poueymirou
JOSEPH SAVERI LAW FIRM, LLP
601 California Street, Suite 1000
San Francisco, CA 94108
Email: jsaveri@saverilawfirm.com
        czirpoli@saverilawfirm.com
        cyoung@saverilawfirm.com
        hbenon@saverilawfirm.com
        lkessler@saverilawfirm.com
        acera@saverilawfirm.com
        mpoueymirou@saverilawfirm.com

Matthew Butterick
MATTHEW BUTTERICK,
ATTORNEY AT LAW
1920 Hillhurst Avenue, #406
Los Angeles, CA 90027
Email: mb@butericklaw.com

*Attorneys for Individual and Representative
Plaintiffs and the Proposed Class*

Bryan L. Clobes (*pro hac vice*)
Alexander J. Sweatman (*pro hac vice*)
Mohammed Rathur
CAFFERTY CLOBES MERIWETHER
& SPRENGEL LLP
135 South LaSalle Street, Suite 3210
Chicago, IL 60603
Email: bclobes@caffertyclobes.com
        asweatman@caffertyclobes.com
        mrathur@caffertyclobes.com

*Attorneys for Individual and Representative
Plaintiffs and the Proposed Class*

Joshua I. Schiller
Maxwell Vaughn Pritt
BOIES SCHILLER FLEXNER
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
Email: jischiller@bsfllp.com
        mpritt@bsfllp.com

David Boies (*pro hac vice*)
BOIES SCHILLER FLEXNER
333 Main Street
Armonk, NY 10504
Email: dboies@bsfllp.com

Jesse Panuccio (*pro hac vice*)
BOIES SCHILLER FLEXNER
1401 New York Ave. NW
Washington, DC 20005
Email: jpanuccio@bsfllp.com

*Attorneys for Individual and Representative
Plaintiffs and the Proposed Class*

Brian O'Mara
DiCELLO LEVITT LLP
4747 Executive Drive, Suite 240
San Diego, CA 92121
Email: BrianO@dicellolevitt.com

Amy Keller (*pro hac vice*)
James A. Ulwick (*pro hac vice*)
Nada Djordjevic (*pro hac vice*)
DiCELLO LEVITT LLP
10 North Dearborn St., Sixth Floor
Chicago, IL 60602
Email: akeller@dicellolevitt.com
        julwick@dicellolevitt.com
        ndjordjevic@dicellolevitt.com

David A. Straite (*pro hac vice*)
DiCELLO LEVITT LLP
485 Lexington Avenue, Suite 1001
New York, NY 10017
Email: dstraite@dicellolevitt.com

Attorneys for Plaintiff Lysa TerKeurst

1   Elizabeth J. Cabraser
2   Daniel M. Hutchinson
    Reilly T. Stoler
3   LIEFF CABRASER HEIMANN &
    BERNSTEIN, LLP
4   275 Battery Street, 29th Floor
    San Francisco, CA 94111-3339
5   Telephone: (415) 956-1000
    Email: ecabraser@lchb.com
6   dhutchinson@lchb.com
            rstoler@lchb.com
7
8   Rachel Geman
    LIEFF CABRASER HEIMANN &
9   BERNSTEIN, LLP
    250 Hudson Street, 8th Floor
10 New York, New York 10013-1413
    Telephone: (212) 355-9500
11 Email: rgeman@lchb.com
12
13 *Attorneys for Plaintiff Christopher Farnsworth*
    and Representative Plaintiffs and the Proposed
14 *Class*

Nancy Evelyn Wolff
COWAN DEBAETS ABRAHAMS &
SHEPPARD LLP
9454 Wilshire Boulevard, Suite 901
Beverly Hills, CA 90212
Telephone: (310) 340-6334
Fax: (310) 492-4394
Email: NWolff@cdas.com

Scott J. Sholder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS &
SHEPPARD LLP
60 Broad Street, 30th Floor
New York, New York 10004
Telephone: (212) 974-7474
Email: ssholder@cdas.com
         ccole@cdas.com

Attorneys for Plaintiff Christopher Farnsworth
and Representative Plaintiffs and the Proposed
Class

15

16

17

18

19

20

21

22

23

24

25

26

27

28

COOLEY LLP
ATTORNEYS AT LAW

                                    2

                                              PROOF OF SERVICE
                                              3:23-CV-03417-VC

COOLEY LLP
BOBBY GHAJAR (198719)
(bghajar@cooley.com)
COLETTE GHAZARIAN (322235)
(cghazarian@cooley.com)
1333 2nd Street, Suite 400
Santa Monica, California 90401
Telephone: (310) 883-6400

MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
JUDD LAUTER (290945)
(jlauter@cooley.com)
ELIZABETH L. STAMESHKIN (260865)
(lstameshkin@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304-1130
Telephone: (650) 843-5000

LEX LUMINA PLLC
MARK A. LEMLEY (155830)
(mlemley@lex-lumina.com)
745 Fifth Avenue, Suite 500
New York, NY 10151
Telephone: (646) 898-2055

CLEARY GOTTLIEB STEEN & HAMILTON LLP
ANGELA L. DUNNING (212047)
(adunning@cgsh.com)
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4131

Counsel for Defendant Meta Platforms, Inc.

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

RICHARD KADREY, et al.,
 	
    Individual and Representative Plaintiffs,
 	
    v.
 	
META PLATFORMS, INC., a Delaware corporation;
 	
                              Defendant.
	Lead Case No. 3:23-cv-03417-VC
	Related Case No. 4:23-cv-06663
	
	DEFENDANT META PLATFORMS, INC.'S FIRST SUPPLEMENTAL RESPONSES AND OBJECTIONS TO PLAINTIFFS' THIRD SET OF INTERROGATORIES

COOLEY LLP
ATTORNEYS AT LAW

1

2

3

4

5

6

PROPOUNDING PARTY:	PLAINTIFFS RICHARD KADREY, SARAH SILVERMAN, CHRISTOPHER GOLDEN, TA-NEHISI COATES, JUNOT DÍAZ, ANDREW SEAN GREER, DAVID HENRY HWANG, MATTHEW KLAM, LAURA LIPPMAN, RACHEL LOUISE SNYDER, JACQUELINE WOODSON, LYSA TERKEURST, AND CHRISTOPHER FARNSWORTH
RESPONDING PARTY:	DEFENDANT META PLATFORMS, INC.
SET NUMBER: ONE THIRD	

7

8

9

10

11

Pursuant to Federal Rule of Civil Procedure 33 and Local Rule 33, Defendant Meta Platforms, Inc. ("Meta") responds as follows to Plaintiffs Richard Kadrey, Sarah Silverman, Christopher Golden, Ta-Nehisi Coates, Junot Díaz, Andrew Sean Greer, David Henry Hwang, Matthew Klam, Laura Lippman, Rachel Louise Snyder, Jacqueline Woodson, Lysa TerKeurst, and Christopher Farnsworth's ("Plaintiffs") Third Set of Interrogatories ("Interrogatories").

12

I.   RESPONSES TO ALL INTERROGATORIES

13

14

15

16

17

18

19

20

1. Meta's responses to these Interrogatories are made to the best of Meta's current employees' present knowledge, information, and belief. Said responses are at all times subject to such additional or different information that discovery or further investigation may disclose and, while based on the present state of Meta's recollection, is subject to such refreshing of recollection, and such additional knowledge of facts, as may result from Meta's further discovery or investigation. Meta reserves the right to make any use of, or to introduce at any hearing and at trial, information and/or documents responsive to these Interrogatories but discovered subsequent to the date of these responses, including any such information or documents obtained in discovery herein.

21

22

23

24

25

26

27

28

2. To the extent that Meta responds to an Interrogatory by stating that Meta will provide information or documents that Meta deems to embody material that is private, business confidential, proprietary, trade secret, or otherwise protected from disclosure pursuant to Federal Rule of Civil Procedure 26(c) and/or Federal Rule of Evidence 501, Meta will only do so subject to the parties' stipulated protective order governing the unauthorized use or disclosure of such information or documents with a designated "CONFIDENTIAL" or "HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY" or "HIGHLY CONFIDENTIAL – SOURCE CODE" distinction (ECF No. 90, the "Protective Order").

1   **3.**   Meta reserves all objections or other questions as to the competency, relevance,

2 materiality, privilege or admissibility as evidence in any subsequent proceeding in or trial of this

3 or any other action for any purpose whatsoever of Meta's responses herein and any document or

4 thing identified or provided in response to the Interrogatories.

5 **II.**   **OBJECTIONS TO INSTRUCTIONS AND DEFINITIONS**

6   Whether or not separately set forth in response to each Interrogatory, Meta makes the

7 following objections to Plaintiffs' Instructions and Definitions:

8   **1.**   Meta objects to all defined terms to the extent that they are not utilized in Plaintiffs'

9 Second Set of Interrogatories.

10   **2.**   Meta objects to the definition of "Communications" to the extent it is inconsistent

11 with and otherwise seeks to circumvent the custodian and search term limits for electronic

12 communications (including emails and other electronic correspondence, and documents attached

13 thereto), as provided in the Stipulated Protocol regarding Electronic Discovery ("ESI

14 Order"). Meta will produce Documents, including Communications, pursuant to the terms of the

15 ESI Order, and any agreement to produce such Documents is explicitly in view of the terms of the

16 ESI Order. To the extent that Meta responds to a Request, including by agreeing to search for

17 relevant, non-privileged communications in Meta's possession, custody, or control, such response

18 is not a representation that any particular custodian or search term is appropriate. Meta expressly

19 reserves the right to object to any custodians and search terms proposed by Plaintiffs.

20   **3.**   Meta objects to the definition of "Describe" as overbroad, unduly burdensome, and

21 disproportionate to the needs of the case to the extent that it purports to require more of Meta than

22 any obligation imposed by law, and would subject Meta to unreasonable and undue burden and

23 expense. Meta also objects to the definition of "Describe" as vague, ambiguous, and unintelligible

24 as to the phrase "desirable to support such statement or make the description complete." Meta

25 construes "Describe" to mean provide a detailed narrative statement or description of specific facts.

26   **4.**   Meta objects to the definition of "Llama Models" as vague and ambiguous as to the

27 terms and phrases "other AI models," "instances," "iterations," "versions," "updates,"

28 "modifications," "original version," "experimental versions," "subsequent versions," and

"refinements to the underlying algorithm, parameters, or architecture," as applied to Llama and "any other AI models developed or in development by Meta." Meta further objects to this definition as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent that it purports to require Meta to produce documents concerning large language models ("LLMs") that were not publicly released and/or were not trained on corpuses of text that allegedly include any of Plaintiffs' allegedly copyrighted works. For the same reason, Meta objects to this definition to the extent that it purports to require Meta to produce documents that are not relevant to any party's claims or defenses. Meta construes "Llama Models" to mean the models within the Llama family of LLMs publicly released by Meta, namely, Llama 1, Llama 2, Code Llama, and Llama 3.

5. Meta objects to the definitions of "Llama 1," "Llama 2," and "Llama 3" as vague and ambiguous as to the undefined terms "precursor models" and "variant models." Meta further objects to these definitions as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent that it purports to require Meta to produce documents or information concerning LLMs that were not publicly released and/or were not trained on corpuses of text that include any of Plaintiffs' allegedly copyrighted works. For the same reason, Meta objects to these definitions to the extent that they purport to require Meta to produce documents or information concerning LLMs that are not relevant to any party's claims or defenses. For purposes of these responses, Meta construes the term "Llama 1" to refer to the LLM released by Meta as Llama on February 24, 2023, the term "Llama 2" to refer to the LLM released by Meta under that name on July 18, 2023, and the term "Llama 3" to refer to the LLMs released by Meta under that name on April 18, 2024, July 23, 2024, and September 25, 2024.

6. Meta objects to the definition of "Meta" as overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent that it purports to require Meta to produce documents outside of its possession, custody, or control. Meta construes "Meta" or "You" to mean Meta Platforms, Inc.

7. Meta objects to the definition of "Relevant Period" as vague, ambiguous, and unintelligible, as it is defined circularly to mean "all times relevant to … the Complaint." Meta construes the Relevant Period to mean January 1, 2022 to the present.

1   **8.**   Meta objects to the definition of "Shadow Dataset(s)" as vague and ambiguous as

2 to the phrasing "the type of databases described in paragraph 37 of the Complaint—databases

3 including but not limited to . . ." and "encompass all versions, updates, augmentations, or

4 modifications of such databases." Without any admission with respect to their content or nature,

5 Meta construes the term "Shadow Dataset(s)" as used in the Requests and the term "Third Party

6 Datasets" synonymously to refer to the third party datasets that are commonly referred to as

7 Books3, Z-Library (aka B-ok), Library Genesis, Bibliotik, Anna's Archive, and The Pile.

8   **9.**   Meta objects to the definition of "SRT" to the extent it characterizes Meta's SRT

9 platform as being used to track "business" issues. Meta construes "SRT" to refer to Meta's SRT

10 platform.

11   **10.**   Meta objects to the definitions of "Train" and "Training" as vague, ambiguous,

12 overbroad, unduly burdensome, and disproportionate to the needs of the case to the extent these

13 terms refer to processes other than pretraining or post-training processes. Meta construes "Train"

14 and "Training" to refer to pretraining and post-training processes.

15   **11.**   Meta objects to Instruction 1 to the extent that it purports to require more of Meta

16 than any obligation imposed by law, and would subject Meta to unreasonable and undue burden

17 and expense. Meta will supplement or amend its responses to these Interrogatories in accordance

18 with Meta's obligations under Rule 26(e).

19   **12.**   Meta objects to Instruction 2, which defines the "Relevant Period" as January 1,

20 2000 to the present. Such definition is overbroad, unduly burdensome, and disproportionate to the

21 needs of the case because it both precedes the existence of Facebook (and therefore Meta) by

22 several years, and the development of the Meta Language Models by decades. For the same reason,

23 the definition of "Relevant Period," as applied to the Interrogatories, would encompass information

24 that is irrelevant to the parties' claims and defenses. The Instruction is also inconsistent with the

25 definition of "Relevant Period" provided on page 3 of the Interrogatories and is therefore vague

26 and ambiguous. Meta construes the Relevant Period to mean January 1, 2022 to the present.

27   **13.**   Meta objects to Instruction 5 (referring to Fed. R. Civ. P. Rule 33(d)) on the ground

28 that it purports to require more of Meta than any obligation imposed by law, and would subject

1   Meta to unreasonable and undue burden and expense.

2   **14.**   Meta objects to Instruction 6 (outlining additional obligations for allegedly

3   incomplete responses) to the extent that it purports to require Meta to investigate information

4   outside of its possession, custody, or control.

5   **15.**   Meta objects to Instruction 7 (outlining additional obligations for any privilege

6   objection or work product objection) on the ground that it purports to require more of Meta than

7   any obligation imposed by law, would subject Meta to unreasonable and undue burden and expense,

8   and would itself require disclosure of information protected by attorney-client privilege and/or

9   attorney work product doctrine.

10   **16.**   Meta objects to Instruction 8 (building in a separate question for each Interrogatory)

11   on the ground that it purports to require more of Meta than any obligation imposed by law, seeks

12   disclosure of information protected by attorney-client privilege and/or attorney work product

13   doctrine, and seeks to circumvent Plaintiffs' interrogatory limit.

14   **17.**   Meta objects to Instruction 10 (purporting to require responses for "all predecessors,

15   successors, subsidiaries … divisions and/or affiliates of Meta"), on the ground that it purports to

16   require more of Meta than any obligation imposed by law, and would subject Meta to unreasonable

17   and undue burden and expense. Meta further objects to Instruction 11 to the extent that it purports

18   to require Meta to investigate information outside of its possession, custody, or control. As such,

19   the Instruction is overly broad as well. Subject to any objections applicable to a particular

20   Interrogatory, Meta will conduct a reasonable, proportionate search for non-privileged, relevant,

21   responsive information within its possession, custody, or control.

22   **18.**   In responding to all Interrogatories, Meta will comply with the requirements of the

23   Federal Rules of Evidence and Federal Rule of Civil Procedure 26.

24   **III.   OBJECTIONS AND RESPONSES TO INDIVIDUAL INTERROGATORIES**

25   **INTERROGATORY NO. 19:**

26   Describe the total revenue You have received and forecast or otherwise expect you will

27   receive from the distribution or use of Llama Models since the start of their development, including

28   by way of any direct sales, licensing fees, or other financial or in-kind benefits.

1   **RESPONSE TO INTERROGATORY NO. 19:**

2   Meta incorporates by reference its objections and definitions above.

3   Meta objects to this Interrogatory as vague and ambiguous as to the term "distribution" and

4   the phrasing "since the start of their development" (as it relates to actual revenues), "direct sales,"

5   and "other financial or in-kind benefits." Meta construes this Interrogatory as seeking information

6   concerning Meta's actual and projected revenue associated with the Llama Models (as construed

7   above).

8   Meta objects to this Interrogatory on the ground that it is overbroad, unduly burdensome,

9   and disproportionate to the needs of the case and seeks information that is not relevant to the parties'

10   claims and defenses.

11   Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

12   33(a)(1).   Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

13   of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

14   total.   Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

15   interrogatories or is in the process of supplementing them. Without leave of Court, Plaintiffs served

16   additional interrogatories exceeding the limit.   By answering interrogatories "19-25," Meta does

17   not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

18   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

19   Protective Order and the ESI , Meta responds as follows:

20   Pursuant to Rule 33(d), Meta refers Plaintiffs to the following documents:

21   Meta_Kadrey_00089020,   Meta_Kadrey_00092364,   and   Meta_Kadrey_00093345,

22   Meta_Kadrey_00093571.

23   In response to Plaintiffs' Fifth Set of Requests for Production of Documents, Meta has

24   conducted an additional search for non-privileged documents in its possession, custody, or control

25   sufficient to show Meta's actual and/or projected revenue associated with the Llama Models (as

26   construed above), if any.   Meta will amend its response to this Interrogatory to identify such

27   documents, if any, after they have been produced.

28   **Meta's First Supplemental Response to Interrogatory No. 19:**

COOLEY LLP
ATTORNEYS AT LAW

META'S FIRST SUPP. RESP & OBJ'S TO
PLTFS' THIRD SET OF ROGS
3:23-CV-03417-VC

This response is designated as Highly Confidential – Attorney's Eyes Only under the Protective Order.

- Meta_Kadrey_00092364 and Meta_Kadrey_00093345 (which are duplicates) indicate that, between November 1, 2023 and July 1, 2024, Meta received ███ ███ ███ ███ ███

- Meta_Kadrey_00093571 indicates that between January and June 2024, Meta received a total of approximately ████████████████████████ ███ ███

In addition to the foregoing, Meta anticipates receiving revenue through it   s revenue sharing agreement with ██████.

Meta further identifies Meta_Kadrey_00156462, which reflects ████████████████ █████████████████████████████████.

INTERROGATORY NO. 20:

Describe any partnerships or collaborations related to the Llama Models You have entered with third parties that have resulted in financial benefits, including the details of such arrangements, such as financial terms, revenue or in-kind benefits received in relation to them.

RESPONSE TO INTERROGATORY NO. 20:

Meta incorporates by reference its objections and definitions above.

Meta objects to this Interrogatory as vague and ambiguous as to the phrases "partnerships or collaborations," "financial benefits," and "in-kind benefits." Meta construes this Request as

1   seeking documents concerning revenues generated by Meta from its agreements with third parties

2   concerning use of the Llama Models (as construed above).

3   Meta objects to this Interrogatory on the ground that it is overbroad, unduly burdensome,

4   and disproportionate to the needs of the case and seeks information that is not relevant to the parties'

5   claims and defenses.

6   Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

7   33(a)(1). Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

8   of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

9   total. Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

10   interrogatories or is in the process of supplementing them.   Without leave of Court, Plaintiffs

11   served additional interrogatories exceeding the limit. By answering interrogatories "19-25," Meta

12   does not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

13   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

14   Protective Order and the ESI , Meta responds as follows:

15   **This response is designated as Confidential under the Protective Order.**

16   Meta has agreements with, respectively, ███████████████████████████

17   ██

18   ███████████████████████████████

19   Pursuant to Rule 33(d), Meta has conducted a reasonable search for and will produce copies

20   of Meta's agreements with Amazon, Google, and Microsoft.

21   **Meta's First Supplemental Response to Interrogatory No. 20:**

22   Meta   identifies   Meta_Kadrey_00212903,   Meta_Kadrey_00213257,

23   Meta_Kadrey_00212679, Meta_Kadrey_00212864, and Meta_Kadrey_00212689 as Meta's

24   collaboration agreement with Amazon; Meta_Kadrey_00212710, Meta_Kadrey_00213055, and

25   Meta_Kadrey_00213435   as   Meta's   collaboration   agreement   with   Google;   and

26   Meta_Kadrey_00213281, Meta_Kadrey_00213303,   as Meta's collaboration agreement with

27   Microsoft.

28   **This response is designated as Confidential under the Protective Order.**

1

2

3

4

5

6

7

8

9

10

11
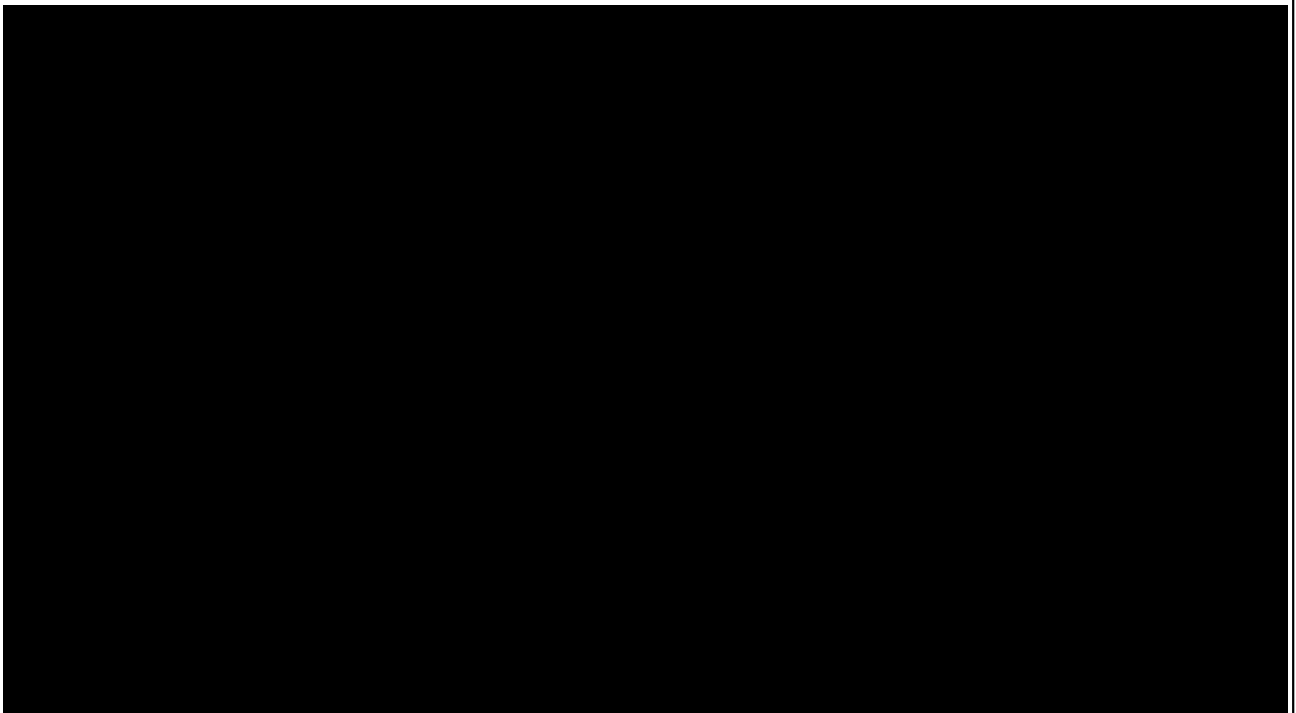
12   **INTERROGATORY NO. 21:**

13   Identify all Person(s) involved in the decision to use Shadow Datasets to train Llama

14   Models.

15   **RESPONSE TO INTERROGATORY NO. 21:**

16   Meta incorporates by reference its objections and definitions above.

17   Meta objects to this Interrogatory as vague and ambiguous as to the phrase "involved in,"

18   which could refer to anyone who had any role in use of such datasets. Meta construes this

19   Interrogatory as seeking the identities of any individuals who had a direct role in the approval of

20   Meta's use of such datasets.

21   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

22   to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

23   in particular to the extent it seeks information concerning individuals whom had no significant

24   involvement in approving use of a particular dataset.

25   Meta objects to this Interrogatory to the extent that it seeks information that is not within

26   Meta's possession, custody, or control.

27   Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

28   33(a)(1).   Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

1 of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

2 total. Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

3 interrogatories or is in the process of supplementing them.   Without leave of Court, Plaintiffs

4 served additional interrogatories exceeding the limit. By answering interrogatories "19-25," Meta

5 does not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

6   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

7 Protective Order, Meta responds as follows:

8   **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

9 **Protective Order.**

10   Meta identifies the following individuals as having had a direct role in approving Meta's

11 use of Third Party Datasets to train the Llama Models (as construed above):

12 • Mike Clark

13 • Lauren Cohen

14 • Ahmad Al-Dahle

15 • Luc Dahlin

16 • Ahuva Goldstand

17 • Praveen Krishnan

18 • Mo Metanat

19 • Jennifer Newstead

20 • Eugene Nho

21 • Mano Paluri

22 • Joelle Pineau

23 • Prashant Ratanchandani

24 • Olena Ripnick-O'Farrell

25 • Steve Satterfield

26 • Ragavan Srinivasan

27 • James Whymark

28 **INTERROGATORY NO. 22:**

1   Describe any efforts You have made to obtain licenses or any similar permissions to use

2 Shadow Datasets, or the works contained therein, to train Llama Models.

3 **RESPONSE TO INTERROGATORY NO. 22:**

4   Meta incorporates by reference its objections and definitions above.

5   Meta objects to this Interrogatory as vague and ambiguous as to the term "works" and the

6 phrases "similar permissions" and "the works contained therein." Meta construes this Interrogatory

7 as seeking information concerning Meta's efforts, if any, to obtain licenses or other consent to use

8 the Third Party Datasets as training data for the Llama Models (as construed above).

9   Meta objects to this Interrogatory to the extent it implicitly assumes that licenses or

10 permission are required for use of the Third Party Datasets to train the Llama Models (as construed

11 above).

12   Meta objects to this Interrogatory because, on its face, it does not exclude legal advice or

13 opinions, which are subject to attorney-client privilege and/or attorney work product doctrine, in

14 particular to the extent that the subject matter of the Interrogatory encompasses internal

15 deliberations between or among Meta attorneys concerning actual or contemplated contract

16 provisions.

17   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

18 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

19 in particular to the extent it seeks information concerning licenses or consent to use works other

20 than literary works as training data for the Llama Models (as construed above).

21   Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

22 33(a)(1). Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

23 of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

24 total. Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

25 interrogatories or is in the process of supplementing them. Without leave of Court, Plaintiffs

26 served additional interrogatories exceeding the limit. By answering interrogatories "19-25," Meta

27 does not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

28   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

1 Protective Order, Meta responds as follows:

2 **This response is designated as Confidential under the Protective Order.**

3 Meta has not sought licenses or other consent, and maintains that it did not need licenses or

4 other consent, to use the Third Party Datasets to train the Llama Models, because such use was fair

5 use.

6 **INTERROGATORY NO. 23:**

7 Identify all sources from which You have obtained Shadow Datasets.

8 **RESPONSE TO INTERROGATORY NO. 23:**

9 Meta incorporates by reference its objections and definitions above.

10 Meta objects to this Interrogatory to the extent that it seeks information that is not within

11 Meta's possession, custody, or control.

12 Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

13 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses,

14 in particular to the extent it seeks information concerning datasets that were not used to train the

15 Llama Models (as construed above) and the identity of every source from which a given dataset

16 may have been obtained at any time.

17 Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

18 33(a)(1).   Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

19 of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

20 total. Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

21 interrogatories or is in the process of supplementing them.   Without leave of Court, Plaintiffs

22 served additional interrogatories exceeding the limit. By answering interrogatories "19-25," Meta

23 does not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

24 Subject to and without waiving the foregoing objections, and pursuant to the terms of the

25 Protective Order, Meta responds as follows:

26 **This response is designated as Highly Confidential – Attorney's Eyes Only under the**

27 **Protective Order.**

28 Following a reasonable investigation, Meta believes that it has obtained the Books3 and

1   The Pile dataset from the website located at <https://the-eye.eu>. With respect to LibGen, Meta

2   obtained links from the LibGen website, <https://libgen.is>, which were used to download dataset

3   files. With respect to Anna's Archive, Meta obtained links from the Anna's Archive website,

4   <https://annas-archive.org>, which were used to download dataset files.

5   **INTERROGATORY NO. 24:**

6   Identify all individuals, including current and former employees, who work or have worked

7   on securing licenses for data or material You have used, have planned to use, or plan to use to train

8   Llama Models.

9   **RESPONSE TO INTERROGATORY NO. 24:**

10   Meta incorporates by reference its objections and definitions above.

11   Meta objects to this Interrogatory as vague and ambiguous as to the phrasing "who worked

12 on securing," which could encompass any number of individuals who lack material knowledge of

13 Meta's data licensing efforts. Meta also objects to the terms "data" and "material" as vague,

14 ambiguous, indefinite, overbroad, and disproportionate to the needs of the case, in particular to the

15 extent it refers to training data other than text. Meta construes this Interrogatory as seeking

16 information concerning the identities of Meta employees who were directly involved in and had

17 responsibility for securing licenses, if any, for text data to train the Llama Models (as construed

18 above).

19   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

20 to the needs of the case and seeks information that is not relevant to the parties' claims and defenses.

21   Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

22 33(a)(1). Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

23 of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

24 total. Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

25 interrogatories or is in the process of supplementing them. Without leave of Court, Plaintiffs

26 served additional interrogatories exceeding the limit. By answering interrogatories "19-25," Meta

27 does not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

28   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

1   Protective Order, Meta identifies the following Meta employees as having been directly involved

2   in and having responsibility for securing licenses, if any, for text data to train the Llama Models:

3   **This response is designated as Confidential under the Protective Order.**

4   • Elisa Garcia Anzano

5   • Faisal Azhar

6   • Alex Boesenberg

7   • Moya Chen

8   • Sy Choudhury

9   • Juan Docuy

10   • Angela Fan

11   • Didem Foss

12   • Brian Gamido

13   • Anna Grosul

14   • Kenneth Heafield

15   • Amanda Kallet

16   • Melanie Kambadur

17   • Sean Koger

18   • Eugene Nho

19   • Aurelien Rodriguez

20   • Shelly Venus

21   **INTERROGATORY NO. 25:**

22   Identify all Meta employees who communicated with third parties—including Microsoft

23   Corporation, OpenAI, Eleuther AI, Anthropic, Dell Inc., Cloudflare, Inc. and Qualcomm—

24   regarding data used to train Llama Models.

25   **RESPONSE TO INTERROGATORY NO. 25:**

26   Meta incorporates by reference its objections and definitions above.

27   Meta objects to this Interrogatory as vague, ambiguous, indefinite, overbroad, and

28   disproportionate to the needs of the case as to the term "third parties." Meta construes this term in

1   the context of this Interrogatory to mean Microsoft Corporation, OpenAI, Eleuther AI, Anthropic,

2   Dell Inc., Cloudflare, Inc. and Qualcomm.

3   Meta also objects to the term "data" as vague, ambiguous, indefinite, overbroad, and

4   disproportionate to the needs of the case, in particular to the extent it refers to training data other

5   than text. Meta construes this Interrogatory as seeking information concerning the identities of

6   Meta employees who were directly involved in and had responsibility for securing licenses, if

7   any, for text data to train the Llama Models (as construed above). Meta construes the term "data"

8   to mean text data.

9   Meta objects to this Interrogatory as overbroad, unduly burdensome, and disproportionate

10   to the needs of the case and seeks information that is not relevant to the parties' claims and

11   defenses.

12   Meta objects to this Request as improperly seeking discovery on discovery.

13   Meta objects to this Interrogatory as exceeding the limit of 25 Interrogatories under Rule

14   33(a)(1). Meta informed Plaintiffs in connection with its objections to their Plaintiffs' prior sets

15   of Interrogatories that those interrogatories included numerous discrete subparts, far exceeding 25

16   total. Nonetheless, and without waiver of its objections, Meta provided its responses to the prior

17   interrogatories or is in the process of supplementing them.   Without leave of Court, Plaintiffs

18   served additional interrogatories exceeding the limit. By answering interrogatories "19-25," Meta

19   does not waive its rights or objections as to Plaintiffs' violation of Rule 33(a)(1).

20   Subject to and without waiving the foregoing objections, and pursuant to the terms of the

21   Protective Order, Meta responds as follows:

22   Meta is presently unaware of any Meta employees who communicated with Microsoft

23   Corporation, OpenAI, Eleuther AI, Anthropic, Dell Inc., Cloudflare, Inc. and Qualcomm

24   regarding text training data used to train the Llama Models. Meta is aware that Mr. Dettmers

25   communicated with EleutherAI about training data for a model unrelated to the Llama Models.

26   However, Meta is in the process of completing a reasonable investigation concerning the subject

27   matter of this Interrogatory and will, if necessary, supplement its response in due course.

28   **Meta's First Supplemental Response to Interrogatory No. 25:**

1   **This response is designated as Confidential under the Protective Order.**

2   Following a reasonable investigation, Meta discloses that Amanda Kallet communicated

3   with ██████ concerning a potential partnership between Meta and ████████

4   ████████████████████ This agreement did not materialize.

5

6   Dated: December 13, 2024   COOLEY LLP

7

8   By: */s/ Judd Lauter*
                      Bobby Ghajar

9   Mark Weinstein
                      Kathleen Hartnett

10   Judd Lauter
                     Liz Stameshkin

11   Colette Ghazarian

12   LEX LUMINA PLLC
                     Mark A. Lemley

13   CLEARY GOTTLIEB STEEN &

14   HAMILTON LLP
                     Angela L. Dunning

15   Attorneys for Defendant
                     META PLATFORMS, INC.

16

17

18

19

20

21

22

23

24

25

26

27

28

PROOF OF SERVICE

I am a citizen of the United States and a resident of the State of California. I am employed in Los Angeles County, State of California, in the office of a member of the bar of this Court, at whose direction the service was made. I am over the age of eighteen years, and not a party to the within action. My business address is Cooley LLP, 355 South Grand Avenue, Suite 900, Los Angeles, CA 90071. On the date set forth below I served the documents described below in the manner described below:

- **DEFENDANT META PLATFORMS, INC.'S FIRST SUPPLEMENTAL RESPONSES AND OBJECTIONS TO PLAINTIFFS' THIRD SET OF INTERROGATORIES**

☒ (BY ELECTRONIC MAIL) I am personally and readily familiar with the business practice of Cooley LLP for the preparation and processing of documents in portable document format (PDF) for e-mailing, and I caused said documents to be prepared in PDF and then served by electronic mail to the parties listed below.

on the following part(ies) in this action:

Executed on December 13, 2024, at Los Angeles, California.

/s/Jerry Gonzalez

Jerry Gonzalez

SERVICE LIST

1

2   Joseph R. Saveri   Joshua I. Schiller, Esq.
    Cadio Zirpoli   Maxwell Vaughn Pritt, Esq.
3   Christopher K.L. Young   **BOIES SCHILLE FLEXNER**
    Holden Benon   44 Montgomery Street, 41st Floor
4   Louis Andrew Kessler   San Francisco, CA 94104
    Aaron Cera   Email: jischiller@bsfllp.com
5   Margaux Poueymirou   mpritt@bsfllp.com
6   **JOSEPH SAVERI LAW FIRM, LLP**
    601 California Street, Suite 1000   David Boise, Esq. (admitted *pro hac vice*)
7   San Francisco, CA 94108   **BOIES SCHILLER FLEXNER**
    Email: jsaveri@saverilawfirm.com   333 Main Street
8   czirpoli@saverilawfirm.com   Armonk, NY 10504
           cyoung@saverilawfirm.com   Email: dboies@bsfllp.com
9   hbenon@saverilawfirm.com
10   lkessler@saverilawfirm.com   Jesse Panuccio (admitted *pro hac vice*)
           acera@saverilawfirm.com   **BOIES SCHILLER FLEXNER**
11   mpoueymirou@saverilawfirm.com   1401 New York Ave. NW
                                            Washington, DC 20005
12 Matthew Butterick   Email: jpanuccio@bsfllp.com
13 **MATTHEW BUTTERICK,**
    ATTORNEY AT LAW   *Attorneys for Individual and Representative*
14 1920 Hillhurst Avenue, #406   *Plaintiffs and the Proposed Class*
    Los Angeles, CA 90027
15 Email: mb@butancricklaw.com
                                            Brian O'Mara
16 *Attorneys for Individual and Representative*   **DiCELLO LEVITT LLP**
    Plaintiffs and the Proposed Class   4747 Executive Drive, Suite 240
17   San Diego, CA 92121
    Bryan L. Clobes (admitted *pro hac vice*)   Email: BrianO@dicellolevitt.com
18 Alexander J. Sweatman (admitted *pro hac*
    vice)   Amy Keller (admitted *pro hac vice*)
19 Mohammed Rathur   James A. Ulwick (admitted *pro hac vice*)
20 **CAFFERTY CLOBES MERIWETHER**   Nada Djordjevic (admitted *pro hac vice*)
    & SPRENGEL LLP   **DiCELLO LEVITT LLP**
21 135 South LaSalle Street, Suite 3210   10 North Dearborn St., Sixth Floor
    Chicago, IL 60603   Chicago, IL 60602
22 Email: bclobes@caffertyclobes.com   Email: akeller@dicellolevitt.com
23   asweatman@caffertyclobes.com   julwick@dicellolevitt.com
           mrathur@caffertyclobes.com   ndjordjevic@dicellolevitt.com
24
    Attorneys for Individual and Representative   David A. Straite (admitted *pro hac vice*)
25 *Plaintiffs and the Proposed Class*   **DiCELLO LEVITT LLP**
                                            485 Lexington Avenue, Suite 1001
26   New York, NY 10017
                                            Email: dstraite@dicellolevitt.com
27
                                            Attorneys for Plaintiff Lysa TerKeurst
28